## Tampere University

LARI LEHTOVIRTA

# Adverse Reaction to Metal Debris in Patients with Metal-on-Metal Hip Replacements

## Etiology & Pathogenesis

LARI LEHTOVIRTA

# Adverse Reaction to Metal Debris in Patients with Metal-on-Metal Hip Replacements
## *Etiology & Pathogenesis*

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology of Tampere University
for public discussion at Tampere University,
on 8 May 2020, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Coxa Hospital for Joint Replacement
Finland

| | | |
|---|---|---|
| *Responsible supervisor* | Docent Antti Eskelinen<br>Tampere University<br>Finland | |
| *Supervisors* | Docent Aleksi Reito<br>Tampere University<br>Finland | Docent Jyrki Parkkinen<br>Tampere University<br>Finland |
| *Pre-examiners* | Professor Petri Lehenkari<br>University of Oulu<br>Finland | Docent Mirva Söderström<br>University of Turku<br>Finland |
| *Opponent* | Professor Tom Böhling<br>University of Helsinki<br>Finland | |
| *Custos* | Associate Professor Ville Mattila<br>Tampere University<br>Finland | |

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

Human knowledge is never contained in one person.
It grows from the relationships we create between each other and the world,
and still it is never complete.

*Paul Kalanithi*

# ABSTRACT

Third generation Metal-on-Metal (MoM) hip replacements were designed as a durable option, especially for young and active people. Traditional metal-on-polyethylene (MoP) would produce polyethylene wear and lead to osteolysis in these patients. Third generation MoM was introduced as a new, longer lasting bearing couple and high hopes were placed on it. Simulator studies showed encouragingly low rates of bearing wear. As a result, surgeons rapidly adopted MoM hip resurfacing and soon MoM total hip arthroplasty (THA) was introduced. However, evidence from clinical studies to support the use of MoM hip replacements was lacking.

More than one million MoM hip replacements were implanted during the early 2000s. First reports of emerging problems were published in 2006. These described periprosthetic soft-tissue lesions causing pain and implant failure leading to revision surgery. In 2007, the Australian Orthopaedic Association National Joint Registry Annual Report reported higher than expected revision rates for MoM hip resurfacings. Several more case series from hospitals were reported, and follow-up programs to identify patients in need of revision surgery were launched. The term Adverse Reaction to Metal Debris (ARMD) was created to describe the diverse findings seen in failed MoM hips.

Etiopathogenesis of ARMD has been of interest for more than a decade now but remains poorly understood. Failure is seen with both high and low wearing hip implants. High wear is, however, considered to be the primary cause of failure in most patients. Metal wear debris is thought to cause local soft tissue inflammation and necrosis in adjacent tissues. Mainly, three types of tissue responses have been suggested: lymphocytic type IV hypersensitivity mimicking response, which has also been termed Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion (ALVAL). The two other types are foreign-body macrophage response and direct cytotoxic response from metal ions, leading to necrosis. Some evidence suggests that the foreign-body and cytotoxic responses are associated with high implant wear or blood metal ion levels, and the lymphocytic hypersensitivity/ALVAL response to low wear, but contradicting reports exist. Patient susceptibility has also

been suggested as a major contributor to the development of soft tissue lesions and subsequent failure.

The aim of this dissertation is to investigate the etiology and pathogenesis of ARMD. In study I, we analyzed bearing wear, whole blood and synovial fluid metal ion levels in MoM hip resurfacings. We then investigated the possible associations of these with histological findings of the synovial tissue. In study II, metal concentrations in synovial tissue were determined and investigated in relation to histological findings, whole blood and synovial fluid metal ion levels. Hip resurfacings and total hip replacements were compared. In study III, we sought to find subtypes of ARMD using statistical cluster and latent class analyses for histological findings of synovial tissues. Imaging findings and metal ion levels were compared between the observed subtypes. In study IV, we aimed to investigate whether intrinsic, host-related factors affect the pathogenesis of ARMD in bilateral MoM hip replacement patients.

In study I, we found that bearing wear, WB and synovial fluid metal ion levels correlated with the degree of macrophage infiltration and necrosis. Further, WB and synovial fluid metal ion levels correlated with bearing wear volume and rate. In study II, we found that periprosthetic tissue metal concentrations were not associated with histological findings. Patients with MoM total hip arthroplasties evinced more necrosis and lymphocytes than did patients with hip resurfacings. In study III, four different subtypes of ARMD were identified. We found that ALVAL-type response is dualistic in nature – either wear-particle related or more of an immunological hypersensitivity response. Cytotoxic and foreign-body responses were also noted. In study IV, it was observed that bilateral patients evince similar histological and imaging findings on contralateral sides despite markedly different wear volumes between the sides.

Our results therefore suggest that ARMD is not one or two entities but four. Implant wear may lead to cytotoxic, foreign-body or wear-related ALVAL response. Some patients may also develop an ALVAL response in the presence of a low wearing hip replacement. Further, intrinsic host-related factors are likely central in the development of ARMD and may dictate the type of tissue response to a large degree. Extrinsic factors, such as wear volume, whole blood and synovial fluid metal ion levels, are associated with the degree of necrosis and the number of macrophages in the tissues. Periprosthetic tissue metal concentrations were not associated with histological features. Thus, the analysis of periprosthetic metal concentrations does not seem beneficial. As the literature regarding the associations between external factors and histological findings is very discrepant, intrinsic factors may be of more importance and lead to susceptibility to metal

debris. Also, taper debris in total hip arthroplasties is likely more immunogenic and/or cytotoxic compared with bearing wear debris. This finding has significance in terms of non-MoM (such as MoP, metal-on-polyethylene) total hip arthroplasties in addition to MoM total hip arthroplasties. In future, understanding why some patients are more susceptible than others and whether these patients can be identified is of great importance to properly allocate follow-up resources and to time revision surgery optimally.

# TIIVISTELMÄ

Kolmannen sukupolven metalli-metallitekonivel (metal-on-metal, MoM) lonkkaan suunniteltiin kulutuskestävyytensä vuoksi erityisesti nuorille ja aktiivisille ihmisille. Perinteinen metalli-muovi-liukupari oli juuri tässä potilasryhmässä osoittautunut huonoksi vaihtoehdoksi muoviosan kulumisen vuoksi. Muovipartikkelit johtivat osalla potilaista luukudoksen heikkenemiseen ja sen seurauksena tekonivelosien irtoamiseen. Kolmannen sukupolven MoM-tekonivelet esiteltiinkin kulutuskestävänä uutena vaihtoehtona. Alustavissa simulaattoritutkimuksissa kuluma näyttäytyi hyvin vähäisenä. MoM-tekoniveliä alettiin suosia nopeasti ja käyttö levisi myös muihin potilasryhmiin, vaikka kliinisten tutkimusten tuomaa näyttöä näiden nivelten tuloksista tai turvallisuudesta ei ollut.

2000-luvun aikana lonkan MoM-tekoniveliä ehdittiin asentaa maailmanlaajuisesti yli miljoona kappaletta. Ensimmäiset viitteet niihin liittyvistä ongelmista saatiin vuonna 2006, kun julkaistiin potilassarjoja, joissa nivelten ympärillä nähtiin kudostuhoa ja pehmytkudosmassoja. Osalla potilaista oireena ilmeni kipua ja liikerajoitusta. Noin vuotta myöhemmin, 2007, Australian ortopediyhdistys ilmoitti heidän rekisterissään näihin tekoniveliin liittyvän odottamattoman paljon uusintaleikkauksia. Useita raportteja uusintaleikkaukseen johtaneista pehmytkudosreaktioista julkaistiin seuraavina vuosina tieteellisissä lehdissä. Termi metallireaktio (Adverse Reaction to Metal Debris, ARMD) lanseerattiin kuvaamaan MoM-tekoniveliin liittyviä tulehduksellisia pehmytkudosreaktioita. Näissä havaittiin nivelkapselin tulehduksellista paksuuntumista, kudostuhoa ja myös isompia tulehduksellisia pehmytkudosmassoja, pseudotuumoreita. Ongelman todellinen laajuus alkoi valjeta vuonna 2010, kun Britannian terveysvalvontaviranomainen julkaisi MoM-tekoniveliin liittyvän varoituksen, jossa suositeltiin leikattujen potilaiden tiivistä seurantaa ja tarvittaessa uusintaleikkausta. Muutamaa kuukautta myöhemmin yksi suosituimmista MoM-tekonivelmalleista vedettiin pois markkinoilta. Tähän päivään mennessä MoM-tekonivelten käyttö on lakannut lähes kokonaan ja useita eri malleja on vedetty pois markkinoilta.

Metallireaktioiden etiologia ja patogeneesi ovat edelleen pitkälti epäselviä huolimatta kiivaasta, vuosia jatkuneesta tutkimustyöstä. Tekonivelestä kuluessa irtoavia nanometrikokoluokan metallipartikkeleja pidetään suurimpana syynä

tulehduksellisiin reaktioihin ja uusintaleikkauksiin. Osalla potilaista uusintaleikkaukseen kuitenkin joudutaan huolimatta hyvin vähäisestä tekonivelen kulumasta. Onkin esitetty pääasiassa kolmen tyyppistä mekanismia näille pehmytkudosreaktioille: metalliyliherkkyyden aiheuttamaa tyypin IV yliherkkyysreaktiota, metallipartikkeleista aiheutuvaa vierasesinereaktiota sekä suoraa metalli-ioneista johtuvaa kudostuhoa aiheuttavaa reaktiota. Tyypin IV reaktiossa keskeisimpinä soluina ovat lymfosyytit ja kudoksissa nähdään nekroosia. Vierasesinereaktiossa taasen makrofagit ovat keskeisimpiä tulehdussoluja eikä nekroosia yleensä nähdä. Suorassa sytotoksisessa reaktiossa metalli-ionit johtavat solujen kuolemaan. Samalla kudoksiin kertyy makrofageja poistamaan kuolleita soluja ja syntyy itseään ylläpitävä tulehdusreaktio. Osassa tutkimuksissa tekonivelten matala kuluma on yhdistynyt tyypin IV yliherkkyysreaktiota muistuttavaan kudosten histologiaan (nimetty ALVAL-reaktioksi, Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion) ja korkea kuluma vierasesinereaktioon tai sytotoksiseen reaktioon. Myös päinvastaisia tuloksia on kuitenkin esitetty ja tutkimuskirjallisuus on monelta osin varsin ristiriitaista. On myös ehdotettu, että potilaiden välillä olisi yksilöllisiä eroja siinä, miten herkästi kudokset reagoivat metallipartikkeleille. Tutkimusnäyttöä tästä ei kuitenkaan ole.

Tässä väitöskirjassa pyrimme tutkimaan metallireaktioiden etiologiaa ja patogeneettisiä mekanismeja. Aineiston muodostivat potilaat, joilta oli Tekonivelsairaala Coxassa uusintaleikattu MoM-tekonivel ja vaihdettu se toisen tyyppiseen tekoniveleen. Keskeisenä tutkimusmetodina oli nivelkapselikudosten histologisten näytteiden analysointi, joka antaa epäsuoraa tietoa tekoniveltä ympäröivien kudosten tilanteesta, tulehdusreaktioista ja niiden syistä. Histologisista analyyseistä saatua tietoa yhdistimme kliinisiin potilastietoihin sekä tekonivelen kulumasta kertoviin mittareihin. Ensimmäisessä osatyössä analysoimme liukuparin kulumaa sekä veren ja nivelnesteen metalli-ionien määrää. Näiden mahdollista yhteyttä nivelkapselin kudoksen histologiaan selvitettiin. Toisessa osatyössä määritimme eri metallien pitoisuuden nivelkapselikudoksessa ja tutkimme sen yhteyttä kudosten histologisiin löydöksiin sekä veren ja nivelnesteen metalli-ionipitoisuuteen. Pinnoitetekonivelten ja kokotekonivelten välisiä mahdollisia eroja verrattiin. Kolmannessa osatyössä pyrimme löytämään mahdollisia piileviä metallireaktion alatyyppejä. Pyrimme ryhmittelemään potilaita tilastomatemaattisesti niiden nivelkapselikudosten histologisten ominaisuuksien perusteella samankaltaisiin ryhmiin. Kuvantamislöydöksiä ja veren metalli-ionipitoisuuksia verrattiin eri ryhmien välillä. Neljännessä osatyössä selvitimme, vaikuttavatko nk. sisäiset tekijät eli potilaskohtaiset erot metallireaktion

kehittymiseen. Tutkimusaineistona oli potilaita, joilta oli molemmista lonkista uusintaleikattu MoM-tekonivel.

Ensimmäisessä osatyössä havaitsimme, että pinnoitetekonivelten kuluma sekä veren ja nivelnesteen metalli-ionipitoisuudet korreloivat kudosten makrofagien määrän ja kudostuhon asteen kanssa. Sen lisäksi veren ja nivelnesteen metalli-ionipitoisuudet olivat yhteydessä tekonivelen liukuparin kuluman määrään. Toisessa osatyössä odotuksista poiketen havaitsimme, että kudosten metallipitoisuus ei ollut yhteydessä kudosten histologisiin löydöksiin. Potilailla, joilla oli uusintaleikattu MoM-kokotekonivel, oli enemmän kudostuhoa ja lymfosyytteja verrattuna pinnoitetekonivelpotilaisiin. Kolmannessa osatyössä löysimme neljä erilaista metallireaktion alatyyppiä. Havaitsimme, että aiemmin julkaistu ALVAL / tyypin IV yliherkkyysreaktio jakautuu mitä luultavimmin kahteen hieman erilaiseen reaktioon – immunologiseen, yliherkkyysmekanismilla syntyvään ja toisaalta metallipartikkeleiden aiheuttamaan. Kahtena muuna alatyyppinä oli metallipartikkeleiden aiheuttama solutuhoon johtava sytotoksinen reaktio ja metallipartikkeleiden aiheuttama vierasesinereaktio. Neljännessä osatyössä näimme, että potilaiden molemmissa lonkissa oli pitkälti samanlaiset histologiset ja kuvantamislöydökset huolimatta isoistakin eroista kulumassa puolien välillä.

Tulostemme perusteella näyttäisi siltä, että ARMD voidaan jakaa neljään eri alatyyppiin. Kuluma voi johtaa sytotoksiseen, vierasesinereaktioon tai ALVAL-reaktioon. Toisaalta osalla potilaista ALVAL voi kehittyä huolimatta hyvin vähäisestä kulumasta. Näiden lisäksi yksilölliset erot potilaiden välillä todennäköisesti ovat keskeisessä roolissa. Nämä saattavat määrittää potilaan herkkyyden metallipartikkeleille ja toisaalta siitä seuraavan kudosreaktion tyypin. Ulkoiset tekijät, kuten nivelen kuluma sekä veren ja nivelnesteen metalli-ionipitoisuudet ovat yhteydessä nekroosin ja makrofagien määrään kudoksessa, mikä tukee aikaisempia havaintoja ja sopii sytotoksisen sekä vierasesinereaktion taudinkuvaan. Tutkimuksissamme tekonivelen liukuparin kuluma myös korreloi veren metalli-ionien pitoisuuteen. Tietoa veren metalli-ionipitoisuudesta voidaan potilastyössä hyödyntää tekonivelen kuluman määrän sekä haitallisen metallireaktion todennäköisyyden arvioinnissa. Mielenkiintoista kyllä, kudosten metallipitoisuus ei odotuksistamme poiketen ollut yhteydessä histologisiin löydöksiin. Kirjallisuus eri ulkoisten tekijöiden yhteydestä kudosreaktion tyyppiin on varsin ristiriitaista. Ehdotamme, että taustalla on ainakin osasyynä potilaskohtaisten, yksilöllisten tekijöiden merkitys patogeneesissä

Tutkimusten potilasaineistot ja menetelmät ovat myös olleet varsin kirjavia, mikä vaikeuttaa tulosten vertailua. Löydöstemme perusteella kartioliitoksesta irtoava metalli on todennäköisesti haitallisempaa kuin liukupinnoista irtoava metalli.

Löydös on merkityksellinen, sillä samantyyppisiä metallisia kartioliitoksia on monessa muussa edelleen käytössä olevassa tekonivelmallissa, esimerkiksi metalli-muovi liukuparisissa. Tulevaisuudessa olisi tärkeää pyrkiä ymmärtämään, miksi osa potilaista näyttäisi olevan herkempiä metallireaktion kehittymiselle tai vaikea-asteisille kudosmuutoksille. Näiden potilaiden aikainen tunnistaminen seurannassa on keskeistä, jotta uusintaleikkaus voidaan tehdä riittävän varhaisessa vaiheessa ja toisaalta, jotta turhilta uusintaleikkauksilta vältyttäisiin.

# CONTENTS

## List of Figures

## List of Tables

# ABBREVIATIONS

| | |
|---|---|
| AIC | Akaikes Information Criteria |
| ALVAL | Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion |
| ALTR | Adverse Local Tissue Reaction |
| AOANJRR | Australian Orthopaedic Association National Joint Replacement |
| ARMD | Adverse Reaction to Metal Debris |
| ASR | Articular Surface Replacement |
| BHR | Birmingham Hip Resurfacing |
| CMM | Coordinate Measuring Machine |
| CoCrMo | Cobalt-Chromium-Molybdenum |
| CPRD | Contact-Patch-to-Rim-Distance |
| EDTA | EthyleneDiamineTetraacetic Acid |
| HCA | Hierarchical Cluster Analysis |
| HR | Hip Resurfacing |
| IQR | Interquartile range |
| LCA | Latent Class Analysis |
| | |
| LD | Large-Diameter |
| LIRC | London Implant Retrieval Center |
| MACC | Mechanically Assisted Crevice Corrosion |
| MARS-MRI | Metal Artifact Reduction Sequence Magnetic Resonance Imaging |
| MHRA | Medicines and Healthcare products Regulatory Agency |
| MoM | Metal-on-Metal |
| MoP | Metal-on-Polyethylene |
| OHS | Oxford Hip Score |
| SD | Standard deviation |
| SD | Small Diameter |
| SF | Synovial fluid |

| | |
|---|---|
| THA | Total Hip Arthroplasty |
| WB | Whole Blood |
| 1$^{st}$ Gen | First generation |
| 2$^{nd}$ Gen | Second generation |
| 3$^{rd}$ Gen | Third generation |

# ORIGINAL PUBLICATIONS

This dissertation is based on the following original publications referred to in the text by their Roman numerals I to IV.

I        Lehtovirta, L., Reito, A., Parkkinen, J., Hothi, H., Henckel, J., Hart, A., & Eskelinen, A. (2017). Analysis of bearing wear, whole blood and synovial fluid metal ion concentrations and histopathological findings in patients with failed ASR hip resurfacings. BMC Musculoskeletal Disorders, 18(1), 523.

II       Lehtovirta, L., Reito, A., Parkkinen, J., Peräniemi, S., Vepsäläinen, J., & Eskelinen, A. (2018). Association between periprosthetic tissue metal content, whole blood and synovial fluid metal ion levels and histopathological findings in patients with failed metal-on-metal hip replacement. PloS One, 13(5), e0197614.

III     Reito, A., Lehtovirta, L., Parkkinen, J., & Eskelinen, A. (2019). Histopathological patterns seen around failed metal-on-metal hip replacements: Cluster and latent class analysis of patterns of failure. Journal of Biomedical Materials Research Part B: Applied Biomaterials.

IV     Lehtovirta, L., Reito, A., Lainiala, O., Parkkinen, J., Hothi, H., Henckel, J., Hart, A., & Eskelinen, A. (2019). Host-specific factors affect the pathogenesis of adverse reaction to metal debris. BMC Musculoskeletal Disorders, 20(1), 195.

# 1 INTRODUCTION

Hip replacements with metal-on-metal (MoM) bearing surfaces were first developed in the 1930s. However, early attempts at using them did not fare well, and for several decades metal-on-polyethylene (MoP) was the bearing surface of choice. (Gomez and Morcuende 2005). The MoP bearing surface functioned well in elderly patients, but in young and active patients the poor wear resistance of polyethylene proved to be a major drawback (Amstutz and Grigoris 1996). At the same time, MoM hip replacements were being developed simultaneously and these new bearings showed promisingly low wear rates in simulator studies (Anissian et al. 1999, Clarke et al. 2000). Subsequently, MoM hip resurfacings were released to the market and directed at young and active people who had high demands for longevity and wear resistance (Amstutz and Le Duff 2006). Early clinical reports were encouraging (Daniel et al. 2004, Back et al. 2005). The use of MoM bearings was soon extended to new-generation large-diameter (LD) total hip arthroplasty (THA) with hopes of reduced dislocation rates and improved longevity compared with conventional MoP THAs (Singh et al. 2013). Unfortunately, the rapidly growing use of MoM bearings was not justified by proper evidence from clinical trials but rather driven by sheer marketing forces and beliefs (Cohen 2012, Reito et al. 2017). By 2012, it was estimated that more than one million patients worldwide had received a MoM hip replacement (Lombardi Jr et al. 2012).

A few years after the large-scale launch of the new-generation MoM hip resurfacings and THAs, the first alarming reports were published. These reports described painful soft-tissue lesions around the joint, which eventually led to revision surgeries (Boardman et al. 2006, Pandit et al. 2008a, Toms et al. 2008). Macroscopically and microscopically, these lesions were very heterogenous and an umbrella term, Adverse Reaction to Metal Debris (ARMD), was launched to describe them as a group (Langton et al. 2010). ARMD is seen with both high and low wearing implants, although high wear is considered more of a risk factor (Kwon et al. 2010, Langton et al. 2010, Matthies et al. 2012, Grammatopoulos et al. 2013).

Histologically, perivascular and diffuse lymphocytic infiltrations and severe necrosis of the periprosthetic tissues were observed in a subset of patients and the response was termed Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion (ALVAL) (Davies et al. 2005, Willert et al. 2005). These findings led to the hypothesis that an adaptive immune response, resembling type IV hypersensitivity response, was the cause of failure in some patients. Later, it was noted that some patients with a failed MoM hip replacement had a macrophage dominated histology instead (Campbell et al. 2010). The authors reported that the ALVAL-type response was associated with low implant wear, suggesting hypersensitivity to wear debris, and the macrophage response was associated with high implant wear, suggesting a response to excess metal wear debris. Thereafter, both supporting and contradicting findings have been published (Grammatopoulos et al. 2013, Lohmann et al. 2013, Paukkeri et al. 2016). Some studies have suggested other responses, such as the cytotoxic effects of metal particles and the immunologically mediated response with tertiary lymphoid organs, in failed MoM hips (Mahendra et al. 2009, Mittal et al. 2013).

Currently, approximately 10 000 patients each year receive a hip replacement in Finland. This number has been steadily growing since the 1980s. (THL 2018). Thus, as the population ages, more demand is placed on the longevity of joint replacements. To achieve a safe and long-lasting hip replacement design, one has to properly understand the mechanisms of implant failure seen with previous bearing materials.

Despite substantial previous and ongoing research efforts, however, the etiology and pathogenesis of ARMD are still poorly understood. Indeed, several fundamental questions remain unanswered. Why do some patients develop destructive lesions despite a well-positioned, low-wearing implant? How is the volume of wear debris generated related to the characteristics of the subsequent tissue response? Are there differences in the susceptibility of individuals to metal wear debris? Do indirect measures of implant wear accurately predict implant wear volume and tissue response? Is debris from modular junctions biologically different than bearing wear debris? Are there possibly several different etiopathogenetic mechanisms of soft-tissue lesions? In this dissertation, we aimed to answer these questions and to further contribute to the complex puzzle of ARMD and its etiopathogenesis.

# 2  REVIEW OF THE LITERATURE

## 2.1  Metal-on-metal bearing in hip arthroplasty

### 2.1.1  History of metal-on-metal hip replacement

The first attempts to treat osteoarthritis with a prosthesis by the French surgeon Pierre Delbet date back to between 1910 and 1920. He performed hemiarthroplasty of the hip using a rubber femoral head to replace the osteoarthritic caput of the femoral bone. During the following three decades, a variety of materials, such as ivory and acryl, were used as a hemiarthroplasty with varying degrees of success. The first total hip arthroplasty (THA), and simultaneously the first metal-on-metal (MoM) hip arthroplasty, was described by Philip Wiles in 1938. He used components made of stainless steel but, unfortunately, with poor results. The first widely used prostheses were developed by Thompson in 1950 and by Böhlman and Moore in 1952. The first person to develop a successful MoM THA was George McKee in the 1950s. Another successful MoM THA was developed by Peter Ring in the 1960s. The MoM hip replacements of this era form the first generation of MoM THA. (Gomez and Morcuende 2005). The use of these MoM implants had ceased by 1970s after John Charnley developed the steel-on-polyethylene THA, which performed better. In the 1980s, interest in MoM arthroplasty grew again when second generation MoM THAs were introduced by Maurice Muller, Bernard Weber and the Sulzer brothers. (Amstutz and Grigoris 1996)

At the same time, first generation MoM hip resurfacings were being developed by Derek McMinn in England and Heinz Wagner in Germany. MoM hip resurfacings were developed based on the hypothesis that the failure of previous hip resurfacing attempts using metal-on-polyethylene (MoP) bearings was due to excessive friction. This led to a myriad of polyethylene wear particles being generated and resulted in osteolysis. The MoP THA had proved to be well-suited for older people, but in young and active people the excessive wear of the

polyethylene was a problem. The proposed benefits of hip resurfacings included preservation of femoral bone, normal joint biomechanics and stability of the joint due to the ability to use large heads. (Amstutz and Grigoris 1996). Early clinical reports of the McMinn resurfacing devices were encouraging (McMinn, Treacy, Lin 1996) and led to the development of the Birmingham hip resurfacing (BHR) (second generation resurfacing) by McMinn (McMinn 2003). Early clinical reports of BHR resurfacings also showed excellent results (Daniel et al. 2004, Back et al. 2005). These MoM resurfacings were targeted at young and active people due to their wear resistance and preservation of bone (Amstutz and Le Duff 2006). As MoM hip resurfacings provided promising results, interest in MoM bearings in THA also grew (third generation) and MoM bearings quickly gained popularity. In 2009, MoM bearings were used in approximately 35% of all THA surgeries performed in the United States (Bozic et al. 2009). In 2012, it was estimated that more than one million patients had received a MoM hip replacement (Lombardi Jr et al. 2012). However, the rapidly increasing use of MoM bearings, especially the Articular Surface Replacement (ASR) (by DePuy Orthopaedics) hip resurfacing and the ASR XL (by DePuy Orthopaedics) THA, was not supported by sufficient evidence from clinical trials (Reito et al. 2017).

Despite the promising early results of MoM bearings, concerns began to be raised some years later when the first reports describing adverse soft tissue reactions around MoM bearings were published (Boardman et al. 2006, Gruber et al. 2007, Pandit et al. 2008a, Toms et al. 2008). Histopathological studies revealed that the majority of the periarticular tissues obtained from patients with failed MoM hip replacements consisted of lymphocyte and macrophage infiltrates and varying amounts of necrosis (Davies et al. 2005, Willert et al. 2005, Pandit et al. 2008b, Mahendra et al. 2009, Campbell et al. 2010). In 2007, the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) stated that the ASR (by DePuy Orthopaedics) and Durom (by Zimmer, Warsaw, IN, USA) hip resurfacings had unexpectedly high revision rates (AOANJRR 2007). The real extent of problem came to light in 2010 when the Medicines and Healthcare products Regulatory Agency (MHRA) in the UK published a medical device alert regarding adverse soft tissue reactions in patients with MoM hips (MHRA 2010). Four months later, in August 2010, DePuy voluntarily recalled the ASR XL and ASR hip resurfacings from the market (DePuy Orthopaedics 2010). It was later shown from internal documents that the marketing and research conducted by DePuy were gravely fraudulent and that the company was aware of the severe problems long before the public and the authorities (Steffen et al. 2018).

Numerous other MoM systems have since been recalled as well: Durom by Zimmer in 2008, R3 by Smith & Nephew in 2012, Rejuvenate and ABG II by Stryker in 2012, and Modular SMF and Modular Readapt Femoral System by Smith & Nephew in 2016 (FDA 2008, 2012b, 2012a, 2016a, 2016b, Smith & Nephew Orthopaedics 2012). As a result of the recalls and high revision rates, the use of MoM bearings in both THA and hip resurfacing surgeries has almost completely ceased (AOANJRR 2017, NJR 2017).

## 2.1.2    Reasons for revision surgery in MoM hips

MoM hip replacements have both traditional modes of implant failure as seen with conventional THA but also failure modes that are more specific to MoM bearings. Traditional failure modes include dislocation, aseptic component loosening, infection, periprosthetic fracture and osteolysis (Carrothers et al. 2010, Reito et al. 2014, Matharu et al. 2016, Seppänen et al. 2016, NJR 2017). However, failures related to metal wear debris resulting from wear of the implant have been of the most concern. Adverse Reaction to Metal Debris (ARMD) is among the most frequent causes of failure in MoM hip replacements. ARMD is an umbrella term and refers to the harmful tissue responses caused by metal wear debris, such as pseudotumors, inflammatory responses, necrosis and metallosis (Langton et al. 2010). In data from the National Joint Registry of England and Wales (NJR), ARMD is the most frequent cause of failure in MoM THA and the second most frequent cause in MoM resurfacings (NJR 2017). In the majority of studies reporting the outcomes of patient cohorts from single centers, ARMD has been the most frequent cause of failure in both THA and hip resurfacings (Ollivere et al. 2009, Langton et al. 2010, 2011b, Reito et al. 2013, Lainiala et al. 2014, Reito et al. 2015a, Matharu et al. 2016, Sidaginamale et al. 2016, Matharu et al. 2017, Lainiala et al. 2019). Moreover, it has been suggested that data from registries underestimate the prevalence of ARMD due to reasons such as the underreporting of MoM failures and the delayed introduction of ARMD as a revision indication (Matharu 2017).

On very rare occasions, the accumulation of cobalt and chromium ions in the bloodstream may lead to systemic consequences. For example, neurotoxicity, cardiomyopathy and thyroid toxicity have all been reported (Bradberry et al. 2014). In their systematic review, Bradberry et al. found 18 case reports of MoM patients with evidence of symptoms caused by the systemic dissemination of metal ions.

Neurological symptoms included peripheral neuropathy, hearing loss and cognitive impairment. Complete or partial resolution of the symptoms were seen in most patients after removal of the metal-containing prostheses.

## 2.2 Wear of the metal-on-metal hip replacement

Implant wear is defined as mechanical action leading to the removal of material (McKellop et al. 2014). Early simulator tests suggested that MoM implants were producing at least an order of magnitude less wear in vitro compared with MoP hip implants (Anissian et al. 1999, Clarke et al. 2000, Goldsmith et al. 2000). However, at the time, it was not properly understood that the biological response to metal particles and ions was significantly different from polyethylene particles, possibly influenced by the smaller particle size, higher number of particles and composition of these particles (Catelas et al. 2011). There are multiple different mechanisms that can produce wear in MoM hip implants, either alone or in conjunction (McKellop et al. 2014). Further, metallic debris can also result from corrosion of the implant, which is not implant wear by definition (McKellop et al. 2014, Hothi et al. 2017). The source of wear may vary depending on the type of MoM implant. In hip resurfacings, wear is mainly produced in the bearing couple, but small amounts of corrosion metal debris may also be generated at the bone-cup interface as well as at the cup-liner interface in modular cup liners (Langton et al. 2010, Vendittoli et al. 2010, Lord et al. 2011, Hothi et al. 2015). In THA, however, wear (and also corrosion) can take place in the head-neck trunnion, the neck-stem trunnion and the femoral stem in addition to the bearing surfaces (Langton et al. 2012, Cooper et al. 2013, Hothi et al. 2016b, 2016a, Di Laura et al. 2018). The source of wear debris has an effect on the particle composition and subsequent tissue response, and is thus of importance (Sidaginamale et al. 2016, Di Laura et al. 2017, Xia et al. 2017).

### 2.2.1 Design and metallurgy of hip resurfacing and total hip replacement

MoM hip resurfacing comprises a metallic cap resurfacing the anatomical femoral head and a metallic cup inserted in the acetabulum (Figure 1) (Amstutz and Le Duff 2006). This combination forms the bearing couple and is the origin of bearing wear debris (Anissian et al. 1999, Clarke et al. 2000). In addition, metal may also be released from the back of the acetabular cup component (Jacobs et al. 1998,

Vendittoli et al. 2010). The back of the cup is most often made of porously coated titanium or cobalt-chromium-molybdenum (CoCrMo) alloy to enhance fixation, whereas the bearing surfaces of the cup and femoral head are made of cast or wrought CoCrMo alloy for low friction and maximum durability (Amstutz and Le Duff 2006, Heisel et al. 2009, Liao et al. 2013). The femoral cap is fixated with cement (Amstutz and Le Duff 2006). The proportions of the metals in the CoCrMo alloy vary, but a common alloy (ASTM standard F75) used in modern hip resurfacings contains 58.9–69.5% cobalt, 27.0–30% chromium, 5.0–7.0% molybdenum and other elements in minor amounts (Mn, Si, Ni, Fe and C) (Liao et al. 2013). Thus, cobalt and chromium are present in significantly higher proportions compared with the other metals.

Large-Diameter (LD) MoM THA shares many of its features with MoM hip resurfacing. It has a similar metallic acetabular cup and a metallic femoral head articulating against the cup, forming a bearing couple identical to that of hip resurfacing. In THA, however, the femoral head is attached to a stem, which is inserted into the femoral medullary canal (Figure 2). The anatomical femoral head is resected in this process. The stem can be fixated with or without cement. If cementless fixation is chosen, a porously coated femoral stem is used to allow for bone ingrowth and stable fixation (Siopack and Jergesen 1995, Mellon et al. 2013b). The head and stem are modular components that are intraoperatively impacted together, forming a taper junction or trunnion. In some designs, the neck is also modularly attached to the stem, forming another neck-stem, modular junction. (Krishnan et al. 2013). These designs are frequently called dual-modular or dual-tapered THA (Cooper et al. 2013). The acetabular cup may be a monobloc (in LD THA) or modular (in small-diameter THA) comprising a separate outer cup and inner liner (Mellon et al. 2013b). The backside of the modular liner is prone to wear and corrosion (Gascoyne et al. 2014, Agne et al. 2015, Hothi et al. 2015, Tarity et al. 2017). It is of significance that all modular junctions are susceptible to wear and corrosion (Higgs et al. 2013). Furthermore, the femoral stem alone, in the absence of modularity, may corrode and produce metal debris (Hothi et al. 2016a). Most of the research regarding modular junctions has focused on material loss at the head-neck trunnion (Cooper et al. 2012, Gill et al. 2012, Langton et al. 2012, Matthies et al. 2013b, Hothi et al. 2016b). Despite the potential for material loss at modular junctions, most importantly at the head-neck trunnion, more material is lost from the bearing couple (Langton et al. 2011a, 2016, Hart et al. 2012c, Matthies et al. 2013b, Scholes et al. 2017).

The metallurgy of the MoM THA resembles that of the MoM hip resurfacing with some differences. However, each manufacturer has their own designs that have some unique properties in terms of metallurgy and the manufacturing process. In MoM THA, the articulating surfaces of the acetabular cup and the femoral head are made of CoCrMo alloy similar to that used in hip resurfacing. There are multiple slightly different alloys used. The most commonly used alloy is F75, the exact metallurgy of which is explained in detail in the first paragraph. Wear and corrosion resistance are the primary reasons for the use of CoCrMo alloy. The outer surface of the acetabular cup is often made of porously coated titanium as in hip resurfacings. The difference with hip resurfacing comes from the femoral stem component. (ASM International 2003). In some designs, CoCrMo is used in the stem, whereas most designs use titanium-based alloys (Krishnan et al. 2013). A popular titanium alloy is composed of 90% titanium, 6% aluminum and 4% vanadium. The advantages of titanium include high biocompatibility due to oxidation of aluminum which forms a passivation layer to the surface. The disadvantages include high potential for wear. As a result, titanium-based alloys are not used in bearing surfaces. (ASM International 2003).

**Figure 1.** Metal-on-metal hip resurfacing components (Biomet Recap). On the left is the acetabular cup and on the right is the femoral resurfacing head, which is hollow from the inside.

**Figure 2.** Metal-on-metal total hip arthroplasty components (DePuy Summit stem, DePuy ASR head and cup). On the left is the acetabular cup and on the right is the modular femoral head attached to the stem.

## 2.2.2 Bearing wear

### 2.2.2.1 Wear modes and wear mechanisms

Bearing wear originates from the bearing couple. Wear modes define the mechanical conditions under which the implant is operating, and four separate wear modes have been defined. Wear mechanisms, on the other hand, are the processes that generate wear at the microscopic level. The wear mode of a well-functioning MoM hip bearing is multidirectional sliding wear (mode 1). Sliding occurs between the head and cup. (Pourzal et al. 2013). Wear modes 2, 3 and 4 describe unintended conditions. In wear mode 2, contact is made between the bearing and non-bearing surfaces. An example of this would be contact between the femoral head and the rim of the acetabular cup. Wear mode 3 refers to a similar condition as in wear mode 1 with the addition of third-body particles between the articulating surfaces. These particles may originate from bone or implant surfaces. Wear mode 4 is defined as contact between two non-bearing surfaces. For instance, contact between the femoral stem and the acetabular rim (impingement). Overall, wear modes are not exclusive to each other and several modes may be present at the same time. (McKellop et al. 2014). Wear mechanisms are the processes that produce damage to the surfaces. Similar to wear modes, there are four distinct wear mechanisms of concern in MoM hip replacements. The first one is adhesive wear, which results from local bonding between two surfaces. As motion occurs, local bonding forces a segment of one surface to break loose, possibly resulting in pits. These loose segments may further act as third-body particles. The second wear mechanism is abrasive wear. Asperities on one surface or third-body particles lead to cutting and plowing, resulting in scratches of a diverse magnitude. The third one is surface fatigue in which cracks on the surface occur. These cracks may also produce loose fragments. (McKellop et al. 2014). The fourth wear mechanism is tribochemical wear. In MoM hip replacements, a tribochemical layer or film is formed on the articulating surfaces when implants are in situ. This layer consists of metallo-organic compounds produced by the mechanical mixing of synovial fluid proteins and the metallic material of the surface layer. The layer modifies the wear behavior of the underlying material and is likely beneficial in reducing adhesion and abrasive wear. However, the tribochemical layer undergoes continuous remodeling – the removal and formation of material – defined as tribochemical wear. In conclusion, wear modes define the

acting wear mechanisms, which may work in conjunction, resulting in material loss. (Wimmer et al. 2009, Pourzal et al. 2013, McKellop et al. 2014).

### 2.2.2.2 Wear in simulator studies – measurement, amount, factors associated with increased wear and properties of the generated metal particles

To better understand the wear behavior of hip implants, joint simulators are used in preclinical studies. The aim of simulators is to reproduce the in vivo conditions of the human hip. The reliability of these simulations, however, depends on how accurately the in vivo conditions can be reproduced. (Affatato et al. 2008). Wear of the implants is measured through gravimetric means and wear volume is further calculated using the weight and density of the material lost (Bills et al. 2005). Several simulator studies on MoM implants were performed at the turn of the century. These studies enhanced our understanding of the wear behavior, amount of wear and the factors associated with increased wear. A study by Clarke et al. showed that the wear of MoM implants is biphasic – consisting of an initial high-wear running-in phase that is followed by a steady-state phase with significantly lower wear (Clarke et al. 2000). In that study, and also in other simulator studies, it has been shown that the overall wear of MoM implants is about one to two orders of magnitude lower compared with traditional MoP implants (Anissian et al. 1999, Clarke et al. 2000, Goldsmith et al. 2000). A meta-analysis comprising 56 simulator studies concluded that the mean running-in wear was 2.1 $mm^3/10^6$ cycles, and the mean steady-state wear was 0.4 $mm^3/10^6$ cycles (Kretzer et al. 2009). $10^6$ cycles in a simulator have been compared to one year of prosthetic use by the patient (Anissian et al. 1999). Kretzer et al. analyzed which designs and manufacturing parameters affected wear behavior the most. They found that high clearance between the head and cup increased running-in wear but not steady-state wear. Contrarily, large head diameter led to lower wear, both running-in and steady-state. It has previously been shown that the significance of these geometrical factors is rooted in their direct effect on the lubrication of the implant (Dowson 2006). In this study, Dowson reported that both the clearance and head diameter affect the thickness of the lubricating fluid film between the articulating surfaces. Whereas lower clearance is generally good for minimizing wear through a thicker fluid film, too low a clearance may lead to equatorial contact, which in turn leads to high friction and wear.

Simulator studies have shown that the wear debris generated at the implant surfaces is composed of nanosized particles in the range of 20-60 nm (Firkins et al.

2001, Catelas et al. 2003). Due to the extremely small particle size, the total number of particles generated by MoM implants exceeds that of MoP implants by a factor of two to three despite the significantly lower volumetric wear of MoM implants (Goldsmith et al. 2000, Firkins et al. 2001). Nanosized particles are considered biologically highly active (Pourzal et al. 2013). A simulator study found that these particles were mostly composed of chromium and oxygen, likely chromium oxide, and to a lesser extent CoCrMo alloy (Catelas et al. 2003). It has been further suggested that cobalt is less present in the particles due to its dissolution into ions (Pourzal et al. 2013). To conclude, early simulator studies showed promisingly low wear of MoM implants that was mostly affected by geometrical factors that led to the generation of high numbers of nanosized wear particles.

### 2.2.2.3 Wear in retrieval studies – measurement, amount and factors associated with increased wear

In retrieval studies, the implants retrieved in revision surgery are examined. Retrieval studies are fundamental for understanding the causes of failure (Jacobs and Wimmer 2013, Hart et al. 2015). Retrieved components should be analyzed thoroughly, that is, inspected visually, microscopically, nanoscopically and measured for material loss (Pourzal et al. 2013). Retrieval analysis can then be combined with clinical patient data (for example, age, sex, BMI, follow-up time, allergological tests) imaging data, blood metal ion concentrations and histological analysis of the periprosthetic tissues to further the understanding of the etiology and pathogenesis of implant failure (Hart et al. 2015). Different methods have been used to estimate bearing surface wear, such as linear wear (maximum wear scar depth) and volumetric wear (total volume of the material lost from bearing surfaces) (Lord et al. 2011). Volumetric wear is considered primary as the total amount of material lost from the surface is of the utmost importance (Ilchmann et al. 2008). However, no single standard exists for the volumetric wear measurement of retrieved implants. Bills et al. described a method developed on the basis of an ISO standard for in vitro wear measurement (Bills et al. 2012). The volumetric

wear was measured using a coordinate machine that probed the explanted implant surface and created a geometrical map of the surface. This map was then compared to a reference map of an unworn surface and the volumetric wear calculated using computer software. Similar geometrical methods have also been used in other studies (Morlock et al. 2008, Becker and Dirix 2009, Witzleb et al. 2009, Lord et al. 2011). However, Bills et al. reported significant measurement uncertainties, which make comparisons between studies challenging (Bills et al. 2012).

Dozens of retrieval studies have been performed over the last three decades (Tables 1 and 2). The reported sample sizes have mostly been small. The data are very heterogenous as some studies have only included data for one component, some studies have included data for both components separately, and some studies have combined the data for both components to calculate total wear. Furthermore, the follow-up time in these studies has been variable and various statistics have been used. However, as can be seen from Table 1, the mean/median wear volume in most studies has ranged between 10 and 100 mm³, and the mean/median volumetric wear rates have been between 2 and 20 mm³/year. In most studies, the mean/median linear wear has been between 10 and 100 µm and the linear wear rate between 5 and 20 µm/year (Table 2). Compared to the mean steady-state wear of 0.4 mm³/10⁶ cycles (comparable to a year of prosthesis use) reported in the simulator study meta-analysis by Kretzer et al., it becomes obvious that the real-world wear seen in the retrieval studies is several-fold higher than that reported in preclinical simulator studies (Kretzer et al. 2009). This highlights the importance of retrieval studies to assess the true performance of prostheses. There are no clearly defined boundaries for abnormal versus normal wear, but volumetric wear rates > 1 mm³/year and linear wear rates > 5 µm/year are generally considered abnormal (Hart et al. 2012a, Sidaginamale et al. 2013, Cook et al. 2019). In most retrieval studies, however, the average values reported exceed these values (Tables 1 and 2). It is therefore safe to state that most of the MoM hip replacements studied produce higher than expected and higher than acceptable amounts of wear debris.

Several factors associated with the high wear of implants have been discovered in retrieval studies, and these can be further categorized into implant-, patient- and surgeon-specific factors. Implant-specific factors include clearance, cup arc of cover and femoral head size (Underwood et al. 2011, Matthies et al. 2013a). Certain implant designs are more susceptible to high wear than others, especially the DePuy ASR hip resurfacing and the DePuy ASR XL THA (Ebramzadeh et al. 2011, Underwood et al. 2011). The ASR hip resurfacing has certain design differences compared with the older generation BHR, that is, reduced arc of cover

of the cup, smaller clearance and lower radius of the acetabular rim (Underwood et al. 2011). Low angle of cup cover, small head size and clearance reduce the "contact patch to rim distance" (CPRD) of the cup and predispose the implant to edge-loading, which is a well-established cause of high wear (Morlock et al. 2008, Matthies et al. 2011, 2013a, Underwood et al. 2011). In a study of highly worn ASR acetabular cups, severe edge-loading was present in all components and on average constituted 58% of the total wear volume (Lu and Ebramzadeh 2019). A recognized patient-specific factor related to high wear is motion pattern (Mellon et al. 2013a). Cup positioning, namely inclination and anteversion, are factors defined as surgeon-specific. Suboptimal cup inclination angle is associated with high wear (Morlock et al. 2008, Hart et al. 2011a, Matthies et al. 2011, Cook et al. 2019). Inclination angle has been shown to correlate with CPRD and edge-loading, which serves to explain the abnormal wear (Morlock et al. 2008, Matthies et al. 2011). In a similar manner, cup anteversion is also related to high wear through its effect on CPRD and the risk of edge-loading (Matthies et al. 2013a, Cook et al. 2019). However, edge-loading leading to high wear may also occur in well-positioned implants (Matthies et al. 2011, Hart et al. 2013). A safe zone of 30-50 degrees inclination and 5-25 degrees of anteversion has been shown to reduce the dislocation rate in THA, later termed "Lewinnek's safe zone" (Lewinnek et al. 1978). In relation to MoM hip replacements, it has been shown that inclination angle outside this safe zone is also related to high wear (Hart et al. 2011a, Matthies et al. 2011). The geometrics of surgeon- and implant-specific factors are illustrated in Figure 3.
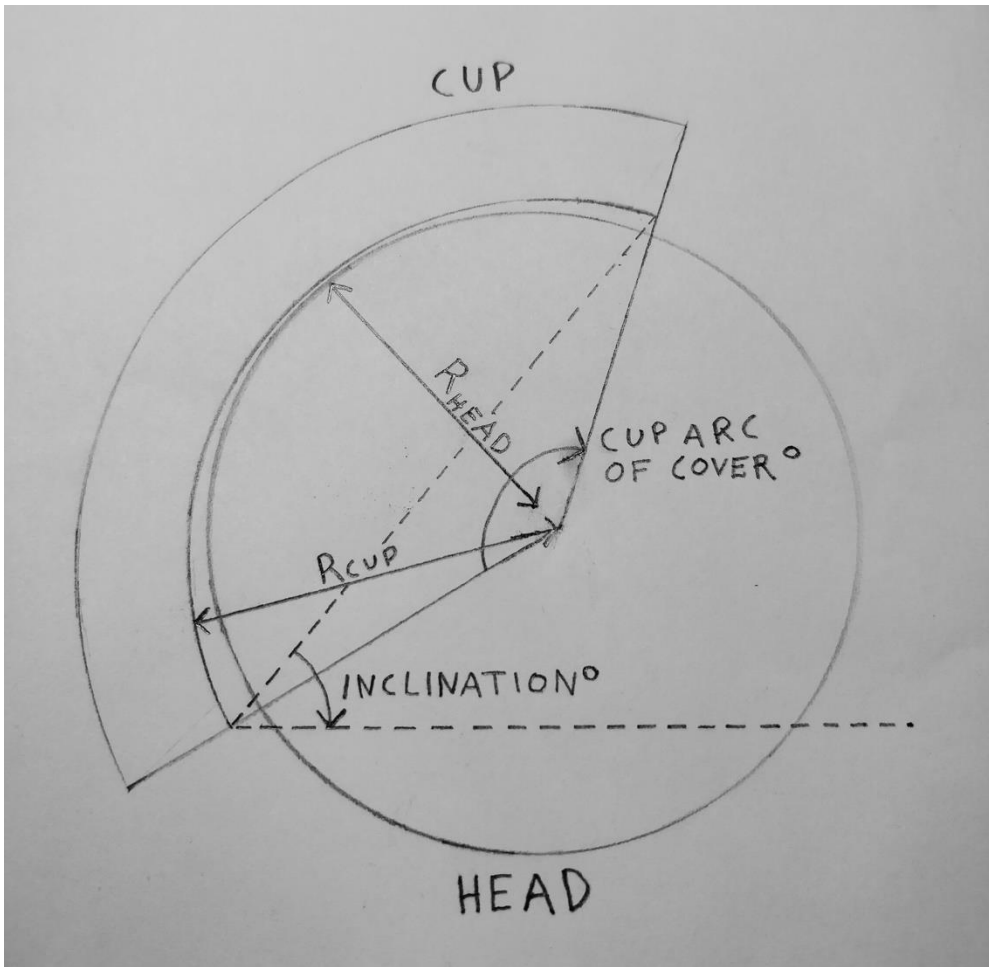
**Figure 3.** A 2-dimensional schematic drawing of the key surgeon- and implant-specific factors related to the wear behavior of hip replacement. Adopted from Matthies et al. 2013a.

*Clearance = $R_{CUP}$ - $R_{HEAD}$

**Table 1.** Studies that have reported the volumetric wear of metal-on-metal hip replacements. THA = Total Hip Arthroplasty, HR = Hip Resurfacing, SD = Small Diameter, LD = Large Diameter, 1st Gen = First generation, 2nd Gen = Second generation, 3rd Gen = Third generation.

| Study | Implant type, subgroup | Component | N | Wear volume (range) $mm^3$ | Volumetric wear rate (range) $mm^3$/year |
|---|---|---|---|---|---|
| McKellop et al. 1996 | 1st Gen, SD THA | Head | 21 | Median 21.7* (1.6-67.3) | Median 2.0 (0.12-11.2) |
| Willert et al. 1996 | 1st Gen, SD THA | Both | 6 | Median 36.87* (0.65-190.4) | Mean 4.96 (0.22-22.36) |
| Milošev et al. 2006 | SD THA | Both | 5 | Mean 1.43 (0.13-3.91) | Mean 0.43 (0.02-1.63) |
| Morlock et al. 2006 | HR, in-situ <100 days | Head | 11 | Mean 0.17 (0-0.47) | - |
| | HR, in-situ 100-200 days | Head | 8 | Mean 1.02 (0-3.61) | - |
| | HR, in-situ >200 days | Head | 7 | Mean 2.76 (0-17.96) | - |
| Bills et al. 2007 | - | Both | 2 | Mean 3.9* (2.2.-6.1) | - |
| Morlock et al. 2008 | HR, edge-load | Both | 14 | - | Median 0.0768 (0.0057-0.1941) |
| | HR, no edge-load | Both | 12 | - | Median 0.009 (0.00-0.0143) |
| Tuke et al. 2008 | 1st Gen SD THA | Both | 5 | Median 35 (28-77) | - |
| Witzleb et al. 2009 | 2nd Gen HR | Head Cup | 8 2 | Median 2.9 (IQR 0.9-6.0) - | Median 2.5 (IQR 0-6.4) Mean 13.4 (4.7-22.1) |
| Bolland et al. 2011 | 3rd Gen LD THA | Both | 17 | - | Mean 1.86 (SD 1.55) |
| Langton et al. 2011a | 3rd Gen LD THA | Both | 51 | - | 1.27 - 24.08 |
| Langton et al. 2011b | 2nd Gen HR | Both | 19 | - | Median 10.83* (3.11 – 95.5) |
| Lord et al. 2011 | 2nd Gen HR | Both | 22 | - | Mean 21.66 (0-51-95.50) |
| Glyn-Jones et al. 2011 | 2nd Gen HR, pseudotumor | Head Cup | 18 18 | Mean 17.4 (± 34.1) Mean 13.4 (± 38.4) | Mean 3.3 (± 5.7) Mean 2.5 (± 6.3) |
| | 2nd Gen HR, no pseudotumor | Head Cup | 18 18 | Mean 2.7 (± 4.0) Mean 1.1 (± 2.4) | Mean 0.8 (± 1.2) Mean 0.4 (± 0.8) |

| | | | | | |
|---|---|---|---|---|---|
| Grammatopoulos et al. 2013 | 2nd Gen HR | Head<br>Cup | 56<br>56 | Mean 15.6 (0-198)<br>Mean 11.3 (0-150) | -<br>- |
| Lohmann et al. 2013 | 2nd Gen SD THA | Both | 28 | - | Mean 0.45 (± 0.31) |
| Pelt et al. 2013 | 3rd Gen LD THA, metallosis | Both | 10 | Mean 94 (± 69) | Mean 16 (± 9) |
| | 3rd Gen LD THA, no metallosis | Both | 7 | Mean 11 (± 10) | Mean 3 (± 2) |
| Matthies et al. 2013a | 3rd Gen LD THA | Head<br>Cup | 110<br>110 | Median 3.44 (0.11–228.30)<br>Median 1.94 (0.06–194.80) | Median 1.31 (0.06–45.66)<br>Median 0.62 (0.04–39.62) |
| Reinisch et al. 2015 | 2nd Gen SD THA | Both | 10 | - | Mean 0.32 (0.22-0.47) |
| Langton et al. 2016 | 3rd Gen LD THA | Both | 47 | - | Median 1.92 (0.23-8.37) |
| Koper et al. 2016 | 3rd Gen LD THA | Head<br>Cup | 9<br>9 | Median 3.2 (0.0-24.4)<br>Median 0.23 (0.0-28.0) | Median 0.94* (0.0-4.8)<br>Median 0.059* (0.0-5.5) |
| Sidaginamale et al. 2016 | 3rd Gen LD THA<br>2nd Gen HR | Both<br>Both | 116<br>83 | -<br>- | Median 2.02 (0.27-68.9)<br>Median 7.35 (0.62-95.5) |
| | 3rd Gen LD THA | Both | 5 | Median 40.9* (30.0-57.3) | Median 6.3 (4.1-7.6) |
| Park et al. 2018 | 2nd Gen HR & 3rd Gen LD THA | Both | 530 | Median 14 (1-636) | Mean 9.0 (0.2-99) |
| Campbell et al. 2018a | 3rd Gen LD THA | Both | 165 | Median 17.5 (-) | - |

\* Calculated from individual measurements given in the study

**Table 2.** Studies investigating the linear wear of metal-on-metal hip replacements. THA = Total Hip Arthroplasty, HR = Hip Resurfacing, SD = Small Diameter, LD = Large Diameter, 1st Gen = First generation, 2nd Gen = Second generation, 3rd Gen = Third generation.

| Study | Implant type & possible subgroups | Component | N | Linear wear (range) $\mu m$ | Linear wear rate (range) $\mu m$/year |
|---|---|---|---|---|---|
| McKellop et al. 1996 | 1st Gen, SD THA | Both | 21 | Median 53* (8-272) | Median 4.9* (0.83-45.3) |
| Schmalzried et al. 1996 | 1st Gen, SD THA | Both | 5 | Median 84* (35-196) | Median 3.8* (1.75-8.9) |
| Willert et al. 1996 | 1st Gen, SD THA | Both | 6 | Median 63.65* (17.1-210) | Median 7.5* (1.67-24.7) |
| Campbell et al. 1999 | 2nd Gen SD THA | Head | 7 | Median 7* (3-19) | - |
|  | 1st Gen HR | Head | 3 | Median 9* (6-32) | - |
| Sieber et al. 1999 | 2nd Gen SD THA | Both | 118 | - | 1st year mean 25 (no range given) |
|  |  |  |  |  | 2nd year mean 5 (no range given) |
| Rieker and Köttig 2002 | 2nd Gen SD THA | Both | 231 | - | 1st year mean 35 (no range given) |
|  |  |  |  |  | After 2nd year 5 (no range given) |
| Böhler et al. 2002 | 2nd Gen SD THA | Both | 6 | Mean 17.3 (6.57-23.16) | Mean 11.1 (2.81-15.44) |
| Reinisch et al. 2003 | 2nd Gen SD THA | Both | 21 | Mean 17 (-) | Mean 7.6 (2.9-12.8) |
| Campbell et al. 2006 | 1st and 2nd Gen HR | Both | 39 | Range <2-164 | - |
| Milošev et al. 2006 | 2nd Gen SD THA | Both | 5 | Mean 31.3 (12.8-52.7) | Mean 6.3 (3.12-9.17) |
| Morlock et al. 2006 | HR, in-situ <100 days | Head | 11 | Mean 5.67 (0-13.00) | - |
|  | HR, in-situ 100-200 days | Head | 8 | Mean 10.50 (4.00-21.00) | - |
|  | HR, in-situ >200 days | Head | 7 | Mean 13.50 (3.00-69.00) | - |
| Bills et al. 2007 | - | Both | 2 | Mean 14.25 (5-25) | - |
|  | 1st Gen SD THA | Both | 5 | Median 84 (30-146) | - |

| Witzleb et al. 2009 | 2nd Gen HR | Head<br>Cup | 8<br>2 | Median 6.9 (IQR 4.5-8.6)<br>- | Median 7.3 (IQR 4.7-9.1)<br>Median 20.35* (9.2-31.5) |
| --- | --- | --- | --- | --- | --- |
| Kwon et al. 2010 | 2nd Gen HR, pseudotumor-group | Head<br>Cup | 9<br>9 | Median 21.05 (2.74-164.80)<br>Median 14.87 (1.93-163.68) | Median 8.1 (2.75-25.4)<br>Median 7.36 (1.61-24.9) |
| | 2nd Gen HR, no-pseudotumor-group | Head<br>Cup | 22<br>22 | Median 4.44 (1.50-8.80)<br>Median 2.51 (0.23-6.04) | Median 1.79 (0.82-4.15)<br>Median 1.28 (0.18-3.33) |
| Matthies et al. 2011 | 2nd Gen HR | Head<br>Cup | 60<br>60 | - | Median 3.50 (0.00-84.70)<br>Median 4.71 (0.00-173.81) |
| | 3rd Gen LD THA | Head<br>Cup | 60<br>60 | - | Median 2.71 (0.00-51.30)<br>Median 3.85 (0.00-119.15) |
| Underwood et al. 2011 | 2nd Gen HR, ASR | Head<br>Cup | 66<br>66 | Mean 13.14 (0.0-315.3)<br>Mean 21.99 (1.3-651.8) | Mean 6.0 (0.0-8.7)<br>Mean 9.2 (0.0-245.6) |
| | 2nd Gen HR, BHR | Head<br>Cup | 64<br>66 | Mean 15.07 (1.5-234.4)<br>Mean 14.9 (2.0-740.4) | Mean 3.5 (0.7-52.4)<br>Mean 4.2 (0.0-153.8) |
| Glyn-Jones et al. 2011 | 2nd Gen HR, pseudotumor | Head<br>Cup | 18<br>18 | Mean 37.8 (± 53.3)<br>Mean 68.4 (± 117.7) | Mean 8.4 (± 8.7)<br>Mean 16.1 (± 21.4) |
| | 2nd Gen HR, no pseudotumor | Head<br>Cup | 18<br>18 | Mean 8.7 (± 10.8)<br>Mean 3.5 (± 5.6) | Mean 2.9 (± 3.9)<br>Mean 1.0 (± 1.5) |
| Hart et al. 2011a | 2nd Gen HR | Head<br>Cup | 45<br>45 | -<br>- | Median 8.7 (IQR 0.0-8.4)<br>Median 5.6 (IQR 3.2-27.2) |
| Ebramzadeh et al. 2011 | 1st and 2nd Gen HR & THA - pseudotumor | Head<br>Cup | -<br>- | -<br>- | Median 8.00 (no range)<br>Median 10.02 (no range) |
| | No pseudotumor | Head<br>Cup | -<br>- | -<br>- | 5.25 (no range)<br>4.15 (no range) |
| Matthies et al. 2012 | 2nd Gen HR, pseudotumor | Head<br>Cup | 72<br>72 | -<br>- | Median 5.3 (0.0–84.1)<br>Median 6.8 (0.0–180.0) |
| | 2nd Gen HR, no pseudotumor | Head<br>Cup | 33<br>33 | -<br>- | Median 2.0 (0.0–62.1)<br>Median 2.2 (0.0–64.3) |

| Hart et al. 2013 | 2nd Gen HR | Cup | 276 | - | Median 4.5 (IQR 0.2 – 20.7) |
|---|---|---|---|---|---|
| Grammatopoulos et al. 2013 | 2nd Gen HR | Head | 56 | Mean 36 (0-283) | |
| | | Cup | 56 | Mean 87 (0-949) | |
| | | Both | 56 | - | Mean 21.8 (0-202) |
| Ebramzadeh et al. 2014 | 2nd Gen HR | Head | 88 | Median 14 (2-316) | - |
| | | Cup | 88 | Median 12 (2-614) | - |
| Takamura et al. 2014 | 2nd Gen HR, ALTR | Head | 7 | - | Median 12.5 (3.6-31.5) |
| | | Cup | 4 | - | Median 23.0 (1.5-36.7) |
| | 2nd Gen HR, no ALTR | Head | 23 | - | Median 1.7 (0.6-7.7) |
| | | Cup | 5 | - | Median 1.48 (0.86-2.53) |
| Matthies et al. 2013a | 2nd Gen HR | Head | 165 | - | Median 4 (0-25) |
| | | Cup | 165 | | Median 5 (0-90) |
| Reinisch et al. 2015 | 2nd Gen SD THA | Both | 10 | - | Mean 1.6 (1.0-2.1) |
| Park et al. 2018 | 3rd Gen LD THA, 2nd Gen HR | Head | 553 | Median 13 (2-649) | - |
| | | Cup | 546 | Median 23 (3-968) | - |

* Calculated from individual measurements given in the study.

## 2.2.3　Trunnion wear and corrosion

It has been stated that metals used as biomaterials should be noble or resistant to corrosion in the human body. The cobalt-chromium alloys used in MoM hip replacements are noted for their high corrosion and wear resistance. (ASM International 2003). However, as previously discussed, all modular junctions in MoM THA have been shown to be prone to both corrosion and wear. These processes are of importance because they lead to the release of metal ions and particles that contribute to adverse host reactions (Cooper et al. 2012). By definition, wear is the removal of material by mechanical action (McKellop et al. 2014). Corrosion, on the other hand, can appear in various forms. Furthermore, corrosion and mechanical wear often interact and can result in further degradation of the metal (ASM International 2003).

Corrosion is an electrochemical process in which two metal surfaces or parts of the same metal surface and their environment interact. The metal surfaces form electrodes and the body fluid between them acts as an electrolyte. The difference in electrochemical potential between the electrodes leads to corrosion and the release of metal ions. (Urish et al. 2019). Implants are protected by the presence of a passive surface film which inhibits corrosion. The disturbance of this film may lead to accelerated corrosion. (Kruger 1979, ASM International 2003).

There are various types of corrosion that include pitting, fretting, crevice, fatigue and galvanic corrosion. In pitting corrosion, there are highly localized areas of corrosion "pits". These may originate, for instance, from defects in the protective surface film. Fretting corrosion occurs in the junctions of two metal surfaces where small micromotion is possible. The micromotion leads to the generation of abrasive particles, mechanical wear and disturbance of the protective film. Crevice corrosion develops when a metal surface is only partially protected from its environment, such as in crevices and junctions. The environment favors corrosion as metal ions accumulate in these areas. When fretting is also present, the process is called mechanically assisted crevice corrosion (MACC). Corrosion fatigue is the fracture of metal resulting from the combined action of corrosion and mechanical forces. Finally, when two dissimilar metals corrode due to the difference in electrochemical potentials, the phenomenon is known as galvanic corrosion. This type of corrosion may occur in the junctions of components made of different metals. (ASM International 2003, Urish et al. 2019).

Of most interest has been the trunnion or taper interface between the neck of the stem and the head in MoM THA. Both wear and corrosion may occur in the trunnion due to micromotion. When these two processes leading to material loss act together, the phenomenon is known as tribocorrosion. The dominating mode of material loss in trunnions is considered to be fretting corrosion or MACC. The taper junction forms a closed crevice where metal ions are robust, and this accelerates the corrosion process. In addition, oxygen is depleted which inhibits the regeneration of the oxidized protective surface film. (Urish et al. 2019).

Corrosion at the trunnion interface can either be qualified by visual inspection or by measuring volumetric material loss (Goldberg et al. 2002, Langton et al. 2012). Visual corrosion scoring has been shown to moderately correlate with actual measured volumetric material loss, the latter being more accurate (Hothi et al. 2014). Although volumetric material loss from the trunnion has been found to be less than bearing wear in the majority of investigated implants, it may also exceed it (Hart et al. 2012c, Langton et al. 2012, Matthies et al. 2013b). Langton et al. observed that a third of implants showed no meaningful surface damage at all, which should be expected from all devices (Langton et al. 2012). It was further observed that the factors associated with higher material loss from the trunnion were large head diameter, varus stems and laterally engaging tapers. Theoretically, all of these factors increase the lever arm and stress encountered at the taper junction. Contrarily to Langton, Matthies et al. found that neither head diameter nor any other design or clinical variable was associated with material loss from the trunnion (Matthies et al. 2013b). It remains unclear therefore which factors accelerate the rate of material loss at the interface.

It is not only the head-neck trunnion that is susceptible to corrosion and metal ion release. Some THA designs include a modular junction between the neck and the stem (dual-modular THA). This junction has been shown to corrode and release metal ions (Krishnan et al. 2013, Di Laura et al. 2018). In addition, the femoral stem as a whole, in regions outside modular junctions, may also corrode (Hothi et al. 2016a).

## 2.2.4 Indirect means for estimating the wear process in MoM hip replacement patients

Wear of MoM prostheses can only be measured after explantation. It is, however, important for clinicians to be able to estimate whether the implant is producing

abnormal amounts of wear as this contributes to the revision decision (Matharu et al. 2018b). Several methods for estimating the amount of wear have been studied. The most established method is the measurement of blood cobalt and chromium ions. Wear and corrosion of the implant produce both insoluble particles and soluble metal ions. The insoluble particles may further oxidize and produce soluble metal ions. (Catelas et al. 2011). These ions diffuse into the bloodstream and can then be used as a surrogate measure of wear. Several studies have shown a strong correlation between blood cobalt and chromium prior to revision surgery and the actual wear volume of the retrieved implants (De Smet et al. 2008, Langton et al. 2011b, Matthies et al. 2013b, Sidaginamale et al. 2013). In a similar manner, synovial fluid metal concentration correlates with wear volume, and a sample can be obtained prior to revision decision (De Smet et al. 2008). The measurement of synovial fluid metal concentrations may, however, be unreliable as metals are present in both soluble and insoluble forms, and most methods only detect soluble metal ions (Davda et al. 2011). Further, synovial fluid sampling is more invasive compared to whole blood sampling and may not offer any additional information.

Several studies have assessed periprosthetic tissues obtained at revision surgery. The disposition of chromium, cobalt, nickel and titanium has been observed (Doorn et al. 1998, Lohmann et al. 2013, Witt et al. 2014, Koper et al. 2016). Witt et al. noted that tissue metal concentrations did not correlate with serum metal levels. Moreover, correlations with implant wear have not been reported. The metal particles observed in the tissues are less than 100 nm in size (Doorn et al. 1998). Several different forms of metals have been observed (Hart et al. 2010, Di Laura et al. 2017, Morrell et al. 2019). The association between metal concentrations in periprosthetic tissues and the wear of the implants has not been studied. Tissue samples, however, can only be obtained at revision surgery. Thus, they are not useful in the follow-up of patients, but they are useful retrospectively for research purposes.

## 2.3    Metals in the human body

### 2.3.1    Metals and health

Certain metals are essential for cellular functioning and human health. These essential metals, which include manganese, iron, copper, zinc, cobalt, molybdenum, chromium and vanadium, belong to trace elements and exist in very limited quantities in the human body. Nutrition is the primary source of these metals. In biologic systems, metals are often bound to proteins, forming metalloproteins, which have essential roles in enzymatic, structural and storage functions. Deficiency of these elements may lead to adverse effects for health and can be reversed with nutrition or supplements. Iron deficiency leading to microcytic anemia is likely the most recognized of these. Contrarily, excess concentrations of these elements may lead to toxicity and organ-specific symptoms. Most of the toxic and beneficial reactions are related to the ability of these metals to participate in redox reactions, either loosing (oxidation) or gaining electrons (reduction). Further, these metals may also bind to proteins and change their conjugation. This may be of benefit or harm. (Fraga 2005). In addition, non-essential metals, such as gold, silver, platinum and lithium, are recognized to have therapeutic effects. Gold has been used to treat severe rheumatoid arthritis, silver has anti-infective properties, platinum has applications in cancer treatment and lithium has proved effective in the treatment of bipolar disorder. (Guo and Sadler 1999).

### 2.3.2    Excess concentrations of metals in the human body

Exposure to too much metal from either the environment or intrinsically from medical devices may lead to excessive concentrations both locally in the affected organ and systemically, observed as elevated blood metal ion concentrations. (Barceloux 1999, Savarino et al. 2002, Lohmann et al. 2013). As discussed earlier, the bearing in MoM hip replacements is composed of CoCrMo alloys. Cobalt and chromium are known to have adverse and toxic effects in high concentrations, but similar data regarding molybdenum are scarce (WHO 1996). The adverse local effects from cobalt and chromium released from MoM hip replacements are discussed in more detail in the next chapter. The systemic effects on these patients

are less well known; however, some rare case reports have been published. These have described neuropathy, cardiomyopathy, hearing loss and cognitive impairment (Bradberry et al. 2014). Cobalt has been considered to be the root cause of these manifestations. In addition, several different cellular mechanisms have been proposed. These include the formation of free radicals, mitochondrial dysfunction, changes in calcium and iron metabolism, changes in erythropoiesis, changes in iodine metabolism and genotoxicity (Paustenbach et al. 2013).

## 2.4    Adverse Reaction to Metal Debris

### 2.4.1    History

The first reports of adverse tissue reactions following the implantation of MoM hip replacements date back to the 1970s. In 1974, Evans et al. reported elevated cobalt and chromium levels combined with necrosis and macrophage infiltration in the periprosthetic tissues of patients implanted with first-generation MoM hip implants (Evans et al. 1974). In addition, they also found that these patients presented evidence of metal sensitivity in skin patch testing. In the same year, Winter found evidence of necrosis and macrophage infiltration associated with the accumulation of cobalt-chromium particles around MoM hips (Winter 1974). In 1977, Willert and Semlitsch published a histological study comparing the findings in the capsular tissues of patients implanted with prostheses of different bearing types (Willert and Semlitsch 1977). In their study, they found that the tissues around MoM hip implants generally displayed macrophages, fibrosis, necrosis and in some cases lymphoplasmocellular infiltrates. In these patients with lymphoplasmocellular infiltrates, hypersensitivity to metal debris was suggested.  In the same year, Brown et al. published their histological findings from around 20 first-generation MoM hips (Brown et al. 1977). All of their samples evinced necrosis and macrophage infiltration, and in some samples lymphocyte infiltration was also noted. As one can observe, adverse response to metal debris from MoM hips was already well described in the 1970s. It has been argued, however, that little attention was paid to these findings during the large-scale reintroduction of MoM hips in the early 2000s (Athanasou 2016).

The problems related to third generation MoM hip implants did not truly surface until relatively recently. Early reports described inflammatory tissue response, necrosis, extra-articular soft-tissue masses, implant loosening and pain (Willert et al. 2005, Boardman et al. 2006, Gruber et al. 2007, Pandit et al. 2008a, Toms et al. 2008). Today, these adverse reactions have resulted in large numbers of patients having to undergo revision surgery. For example, in the Australian Orthopaedic Association National Joint Replacement Registry, a total of 17 345 MoM hip resurfacings had been implanted by 2018. Of these, 1 656 (9.5%) have since undergone revision surgery (AOANJRR 2018). In a recent large cohort of MoM patients, 36% of MoM THAs and 14% of MoM hip resurfacings had undergone revision surgery (Lainiala et al. 2019). It has been estimated that globally at least 80% of all MoM hips remain unrevised (Matharu 2017). Data from our institution is in line with this estimation – 78% of all MoM hip replacements remained unrevised as of 2019 (Lainiala et al. 2019). In total, approximately 1.5 million MoM hip replacements were implanted worldwide and those unrevised are still at risk for failure due to ARMD (Matharu et al. 2018c). Due to the high failure rate caused by adverse reactions, numerous systems have been recalled by their manufacturers (Chapter 2.1.1) and the use of MoM bearings has now almost completely ceased (AOANJRR 2017, NJR 2017).

## 2.4.2 Definition and terminology

Various terms have been used to describe the adverse reactions related to MoM hip replacements. The most widely used has been the term Adverse Reaction to Metal Debris (ARMD), which was first described by Langton et al. in 2010 (Langton et al. 2010). ARMD is an umbrella term and refers to the pathological findings – both micro- and macroscopic – seen in failed MoM hips. These findings include inflammatory tissue responses, soft-tissue masses, necrosis and the macroscopic staining of the tissues, that is, metallosis. Some studies have used the term Adverse Local Tissue Reaction (ALTR) rather than ARMD but both terms can be considered synonyms (Whitehouse et al. 2015, Liow et al. 2016, Xia et al. 2017). The soft-tissue masses, cystic or solid, related to MoM hip implants were described as pseudotumors by Pandit et al. in 2008 and the term has since been widely used (Pandit et al. 2008a). The term Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion (ALVAL) was first defined by Willert et al. in 2005 to describe a

subset of patients with specific histological findings related to the failure of MoM hips (Willert et al. 2005).

The misuse of the terminology described above has been of some concern (Athanasou 2016). For example, ALVAL and ARMD have been used synonymously, although ALVAL refers to a specific histological subset of ARMD. In addition, some studies have used the term pseudotumor as a synonym for ARMD. By definition, all patients with a pseudotumor have ARMD, but not all patients with ARMD have a pseudotumor. Furthermore, the misuse of the terminology in the literature may lead to difficulties when comparing studies.

### 2.4.3    Etiology and risk factors

Most risk factors for ARMD and related failure are similar to the risk factors for all-cause revision (of which most are due to ARMD). However, the different definitions of ARMD between studies makes the evaluation of the risk factors challenging. The reported risk factors for failure specifically related to ARMD and pseudotumor are listed in Table 3. ARMD can be understood as a complication arising from the generation of metal debris from the bearing surface or modular interface. The generation of metal debris leads to local tissue responses and typical clinical presentation (Mahendra et al. 2009, Langton et al. 2010, Haddad et al. 2011). ARMD, in its different forms, has been observed in patients with both high- and low-wearing prostheses (Campbell et al. 2010, Kwon et al. 2010, Langton et al. 2010, Ebramzadeh et al. 2011, Langton et al. 2011b, Matthies et al. 2012, Grammatopoulos et al. 2013, Ebramzadeh et al. 2014). Most failures appear to be related to excess wear, and it is therefore considered to be the most important cause of the development of ARMD (Langton et al. 2010, 2011a, 2011b, Matthies et al. 2011, Takamura et al. 2014). Moreover, high wear also results in high local concentrations of metal debris leading to cellular and immunological cascades, and ultimately manifesting clinically as ARMD (Athanasou 2016). However, ARMD has also been observed in patients with low implant wear. Sensitivity to metal debris, either type IV delayed adaptive immunity response or some other mechanism, has been suggested as a cause of ARMD in these patients (Campbell et al. 2010, Matthies et al. 2012, Park et al. 2018).

Metal ion levels in blood can be used to estimate the in vivo wear state of the implant and the need for revision surgery (De Smet et al. 2008, Langton et al.

2011b, Sidaginamale et al. 2013, Van Der Straeten et al. 2013). High metal ion levels are associated with higher risk for all-cause and ARMD revision when compared with low metal ion levels (Langton et al. 2010, Hart et al. 2011b, Langton et al. 2013b, Hart et al. 2014). However, there is no method for recognizing those who are at risk for revision despite low implant wear and low blood metal ion levels. Metal hypersensitivity mediated by type IV delayed response has been suggested as a cause, but there is no reliable test to predict the individual response to metal debris (Teo and Schalock 2016). It has been reported that women are at higher risk for ARMD and related failure (Glyn-Jones et al. 2009, Murray et al. 2012a, Langton et al. 2013b, Reito et al. 2013, Matharu et al. 2016). In addition, there is some evidence that metal hypersensitivity may be more frequent in women (Ebramzadeh et al. 2011) and previous exposure to metal ions from wearing jewelry has been suggested as a possible cause (Pandit et al. 2008a). On the other hand,  Langton et al. noted that ALVAL-type responses share many histological features with autoimmune diseases, such as rheumatoid arthritis, and these diseases are more frequent in women (Langton et al. 2013b). Thus, it could be that women are genetically more susceptible to ARMD. Patients with bilateral MoM hips have higher risk for failure due to ARMD compared with patients with unilateral MoM hips (Langton et al. 2016). Sensitization to metal debris caused by the first MoM hip was suggested by the authors. In addition to delayed-type hypersensitivity to metal debris, it is possible that the magnitude and type of host response, that is, patient susceptibility, is individually variable overall. This could lead to a difference in the thresholds of metal debris needed between patients to provoke an adverse reaction. Patient susceptibility has been suggested as a possible contributor to the development of ARMD in numerous previous studies (Campbell et al. 2010, Donell et al. 2010, Ebramzadeh et al. 2011, Hart et al. 2012a, Matthies et al. 2012, Ebramzadeh et al. 2014, Athanasou 2016, Langton et al. 2016), but no direct evidence to support this has been presented.

Other known risk factors for ARMD include those factors that affect the wear process of the prosthesis, and these factors can be divided into implant-specific factors and surgical factors. Hip resurfacings have a lower prevalence of ARMD than stemmed THA (Langton et al. 2011a, Reito et al. 2013). Wear debris from the trunnion-interface between the head and stem in THA was suggested by the authors as a contributor to the higher risk of failure due to ARMD. In addition to implant type, implant brand has also been shown to be associated with risk for failure due to ARMD (NJR 2017). The highest failure rates have been observed

with ASR XL and ASR resurfacing devices in data from both registries and clinical trials (Langton et al. 2011b, NJR 2017). It has been suggested that this high failure rate is due to the design of the ASR cups, that is, small arc of cover leading to edge loading and increased wear (Langton et al. 2011b). In some studies utilizing univariable analyses, a small head size in hip resurfacing patients has been identified as a risk factor for ARMD (Glyn-Jones et al. 2009, Ollivere et al. 2009, Langton et al. 2011b, Murray et al. 2012a, Reito et al. 2013), but in analyses controlling for gender, a similar association has not been found (Glyn-Jones et al. 2009, Murray et al. 2012a, Reito et al. 2013, Matharu et al. 2016). Thus, it is likely that the results of the univariable analyses are confounded by the fact that women have smaller components and that female gender is associated with higher risk for ARMD. Conversely, in THA, a large head size has been identified as an independent risk factor for ARMD (Reito et al. 2015a). It has been suggested that increments in head size lead to increases in the lever arm between the trunnion and the head and/or increased frictional torque in bearing, which may translate in to greater micromotion and subsequent wear at the trunnion. This could therefore serve to explain the risk associated with larger head sizes (Langton et al. 2012). Surgical factors associated with increased risk for ARMD include malpositioning of the cup, namely excessive inclination and excessive or insufficient anteversion of the cup (De Haan et al. 2008, Langton et al. 2011b). Excessive inclination leads to increased bearing wear, which is considered to be the root cause of ARMD (Hart et al. 2013).

**Table 3.** The risk factors for revision specifically related to ARMD or pseudotumor in single-center cohorts in patients with metal-on-metal hip replacements.

| Risk factor | Study |
| --- | --- |
| Female gender | Glyn-Jones et al. 2009, Murray et al. 2012b, Reito et al. 2013 |
| Acetabular component malpositioning | De Haan et al. 2008, Grammatopoulos et al. 2010, Langton et al. 2011b |
| High component wear | Kwon et al. 2010, Glyn-Jones et al. 2011 |
| High blood metal ion levels | Langton et al. 2010, 2013b |
| Bilateral hip replacement | Langton et al. 2016 |
| Total hip replacement (vs. hip resurfacing) | Langton et al. 2011a |
| Large head-size in THA | Reito et al. 2015a |
| Age < 40 | Glyn-Jones et al. 2009 |

## 2.4.4 Prevalence of revision surgery resulting from ARMD

The prevalence of revision surgery resulting from ARMD varies widely between studies (Table 4). At our institution, the prevalence has been between 7.4% and 48.8% at five to eight-year follow-up, depending on implant type and brand (Reito et al. 2013, 2015a, Lainiala et al. 2014). The prevalence appears to be higher in patients with THA versus hip resurfacing and in patients with DePuy ASR prostheses versus other devices (Langton et al. 2011a, 2011b, 2013b, 2016, Matharu et al. 2017). In addition to patient, implant and surgeon related factors, Reito et al. have shown that the prevalence of revision for ARMD depends on the level of screening implemented (Reito et al. 2016b). More rigorous screening leads to a higher prevalence of ARMD. Hence, differences in the definition of ARMD, the level of screening and the diagnostic methods used make comparisons between studies challenging. Further, the indications for revision surgery may also vary greatly between centers and individual surgeons (Matharu et al. 2019) Overall, ARMD is the most frequent cause of revision surgery in both HR and THR (Ollivere et al. 2009, Langton et al. 2010, 2011a, Reito et al. 2013, Lainiala et al. 2014, Matharu et al. 2016, 2017, Sidaginamale et al. 2016).

**Table 4.**    Prevalence of revision surgery resulting from ARMD. HR = hip resurfacing, THA = total hip arthroplasty. KM = Kaplan-Meier.

| Study | Brand | Type | Follow-up in years | Diagnosis | Revised for ARMD | Statistics |
|-------|-------|------|--------------------|-----------|------------------|------------|
| Ollivere et al. 2009 | BHR | HR | 5 | Histological | 3.1% | KM |
| Glyn-Jones et al. 2009 | Several | HR | 8 | Histological | 4% | KM |
| Langton et al. 2010 | ASR | HR | 2.92 | Histological | 3.2% | KM |
| | ASR | THA | 3.42 | -"- | 6.0% | KM |
| Langton et al. 2011a | ASR | HR | 6 | Histological/clinical | 25% | KM |
| | | THA | 6 | Histological/clinical | 48.8% | KM |
| Reito et al. 2013 | ASR | HR | 5 | Macroscopic/histological | 25% | Prevalence |
| | | THA | 5 | Histological/clinical | 31% | Prevalence |
| Lainiala et al. 2014 | Pinnacle | THA | 7.5 | Macroscopic/histological | 7.4% | Prevalence |
| Reito et al. 2015a | ASR | THA | 8 | Macroscopic/histological | 43% | KM |
| Langton et al. 2016 | Pinnacle | THA | 9 | Histological/clinical | 16.4% | KM |

## 2.4.5 Clinical presentation

Clinical presentation in ARMD is considered to result from the accumulation of metal debris in periprosthetic tissues, leading to the activation of immunological responses and possible direct cytotoxic effects, further observed as pathophysiological changes, such as inflammation, swelling and necrosis (Figures 4 and 5), which are likely the root cause of clinical manifestations (Mahendra et al. 2009, Langton et al. 2010, Haddad et al. 2011, Athanasou 2016). Patients may present with localized pain, effusion or masses around the hip (Figures 4 and 5), limited range of movement, sensation of pressure or instability, squeaking and numbness of the hip (Pandit et al. 2008a, Langton et al. 2010, Lainiala et al. 2014). In addition, ARMD may be accompanied by gluteal muscle atrophy, which can impair the function of these muscles (Toms et al. 2008, Berber et al. 2015, Reito et al. 2016a). However, patients may be completely asymptomatic despite the inflammation, soft-tissue masses and tissue destruction present in their hips (Kwon et al. 2011, Wynn-Jones et al. 2011, Almousa et al. 2013, Fehring et al. 2014). The natural history of these changes and the possible progression of symptoms is not well understood. Moreover, even less is known about the follow-up and treatment of these asymptomatic patients.
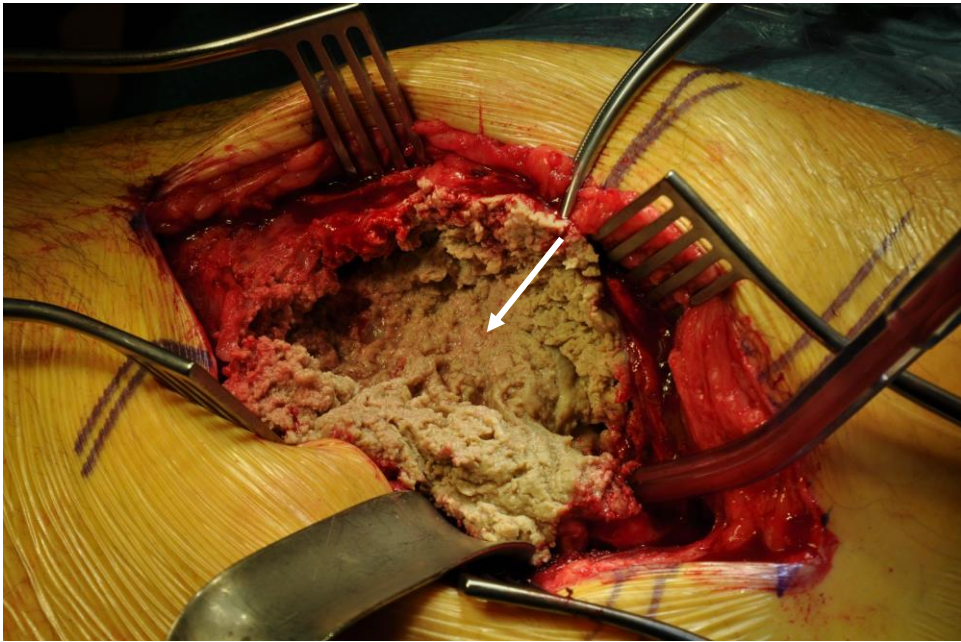
**Figure 4.** A massive necrotic soft-tissue mass (pseudotumor) around the hip joint (arrow).
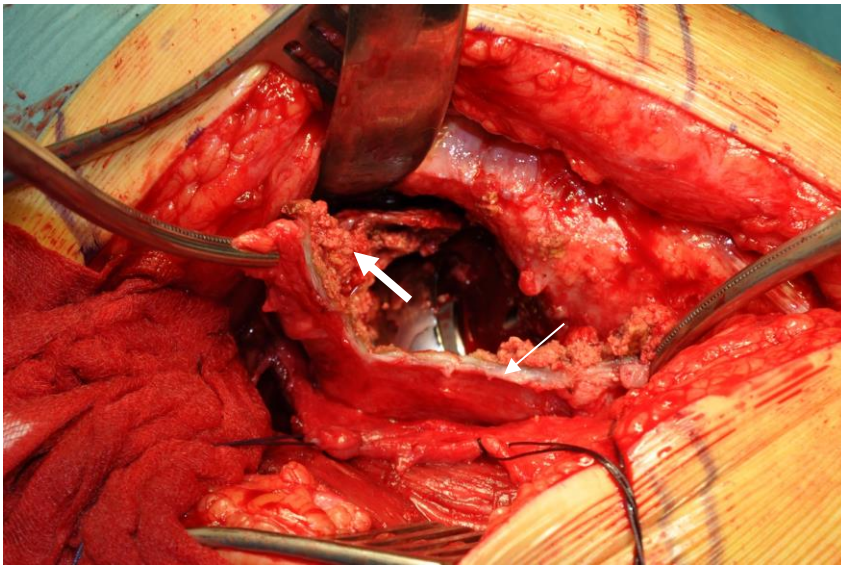


**Figure 5.** A thickened synovial capsule (thin arrow) with necrosis (thick arrow).

## 2.4.6 Surveillance and diagnostics

Patients with MoM hip arthroplasty are at risk for developing ARMD, and therefore in dire need of surveillance. A large body of literature on the subject exists and many authorities worldwide have published their recommendations for follow-up protocols (Health Canada 2012, FDA 2013, MHRA 2017, TGA 2017). Matharu et al., however, have argued that current follow-up protocols are neither evidence-based nor cost effective (Matharu et al. 2015). Most commonly, surveillance includes clinical examination, blood metal ion sampling, imaging and in some cases hip aspiration (Matharu et al. 2018b).

Patients should be evaluated thoroughly for signs and symptoms of ARMD. Symptoms include pain and mechanical sensations, such as clicking, clunking or instability. Signs that can be observed in clinical examination comprise swelling of the hip, soft-tissue masses around the hip area and limping. Patient-reported outcome measures, such as Oxford Hip Score (OHS), may offer supporting information on the severity of patient's symptoms. (Matharu et al. 2018b)

The rationale behind blood metal ion sampling is that in vivo wear is known to correlate with metal ions measured in whole blood and serum (De Smet et al. 2008, Sidaginamale et al. 2013). Furthermore, elevated levels of blood metal ions have been associated with poor implant function and increased risk for implant failure leading to revision surgery (Langton et al. 2010, Hart et al. 2011b, Van Der Straeten et al. 2013, Hart et al. 2014). There is, however, no clear consensus regarding the optimal threshold values of metal ion levels to detect patients with poorly performing MoM implants in need of revision surgery (Matharu et al. 2018b). Therefore, the decision for revision surgery should not be based solely on blood metal ion levels (Hart et al. 2014).

Imaging of the hip is recommended because it provides important information to aid in the decision-making process. Suitable imaging modalities include native radiographs, Metal Artifact Reduction Sequence Magnetic Resonance Imaging (MARS-MRI) and ultrasound. Native radiographs cannot be used to detect soft tissue–related pathologies, but they can identify other abnormalities that are sometimes associated with ARMD. These abnormalities include high acetabular inclination and osteolysis. Furthermore, native radiographs may reveal other causes of implant failure besides ARMD, such as periprosthetic fracture and aseptic loosening of the components. Ultrasound and MARS-MRI are recommended imaging modalities for detecting soft tissue pathologies, such as pseudotumors, severe muscle atrophy and joint effusion, that belong under the umbrella term

ARMD. Both of these imaging modalities have their own advantages and disadvantages. The advantages of ultrasound in comparison to MARS-MRI include its relatively low cost, the short amount of time needed to perform the investigation and fewer contraindications. The advantages of MARS-MRI include not being operator-dependent, better visualization of deeper tissues and the possibility to compare findings to contralateral hip and to retrospectively inspect images. Two classifications have been published for MRI findings and these have been commonly used in later studies: the Imperial classification by Hart et al. and the modified Norwich classification by Anderson et al. (Anderson et al. 2011, Hart et al. 2012d). Both ultrasound and MARS-MRI can detect pseudotumors with sufficient sensitivity and specificity, albeit MARS-MRI has performed slightly better in imaging studies. Hence, MARS-MRI is considered to be the gold standard and is often the first line of imaging. (Matharu et al. 2018b)

Hip aspiration is not frequently used, but in some cases it may aid the diagnostics. For instance, bacterial cultures and cell counts from hip aspirates may help in differential diagnostics between prosthetic joint infection and ARMD. Further, the implant wear of both hips can be estimated separately in bilateral MoM patients using metal content analysis of the hip aspirate. (Matharu et al. 2018b)

In conclusion, no single investigative modality can solely be used in the surveillance of patients with MoM hip implants. Surveillance should include thorough clinical examination, measurement of blood metal ions, native radiographs and cross-sectional imaging with either MARS-MRI or ultrasound. Asymptomatic patients with underlying ARMD present a diagnostic challenge. Moreover, asymptomatic patients may develop pseudotumors with the potential to cause tissue destruction, even in the absence of high blood metal ions (Pandit et al. 2008a, Mahendra et al. 2009, Wynn-Jones et al. 2011, Matthies et al. 2012). It has been suggested that susceptibility to metal debris is the underlying cause in these patients (Willert et al. 2005, Campbell et al. 2010, Matthies et al. 2012). Skin patch testing and lymphocyte transformation tests have been researched as diagnostic means to detect these patients, but evidence does not support their use (Teo and Schalock 2016).

## 2.4.7    Treatment

The heterogeneity of ARMD as a complication and the lack of evidence makes the management and treatment of these patients challenging. Surgeons need to decide whether to continue surveillance or to proceed with revision surgery. There is no good evidence regarding the threshold of revision surgery to achieve optimal results. In some situations, the decision to proceed with revision surgery can be straightforward: symptomatic patient with solid pseudotumor causing damage to adjacent tissues. Contrarily, a patient with no symptoms, no imaging findings and low blood metal ions can safely continue to be surveilled. However, large numbers of patients fall somewhere in between these two scenarios, that is, into a "gray" area. Early studies reporting results of AMRD revisions showed poor outcomes, which led the authorities to suggest early revision. As of today, there is no good quality evidence to support surgeons in the decision-making process; most management guidelines rely on expert opinion (level 5 evidence). (Matharu et al. 2018a).

Due to the heterogeneity of ARMD, revision surgery is often challenging, and surgical strategies need to vary according to the presentation. ARMD with only synovitis and slight metallosis requires a different approach than ARMD with solid pseudotumor, soft-tissue necrosis and muscle atrophy. However, despite the differences in intraoperative presentation, the main objective of revision surgery is to convert the MoM bearing couple into a bearing couple that does not produce metal debris. Suitable bearing couples include MoP, ceramic-on-polyethylene and ceramic-on-ceramic. Depending on the fixation and positioning of the components, either femoral/acetabular or both components can be revised. In THA, revision of only the modular components, such as the femoral head and acetabular liner, may be sufficient. (Matharu et al. 2018a).

ARMD revisions are not without complications. These include infections, dislocation and ARMD recurrence (Matharu et al. 2018a). Early short-term results suggested that the functional outcome of revision is significantly worse compared with MoM revision performed for causes other than ARMD (Grammatopolous et al. 2009). Furthermore, 50% of patients revised for ARMD presented with complications and one third required re-revision. Matharu et al. conducted a systematic review of all studies reporting complication and re-revision rates after ARMD revision (Matharu et al. 2018a). Complication rates were found to range from 4 to 69% and well over 10% in most of the included studies. Re-revision rates ranged from 2 to 44%. The authors noted that the overall outcomes of ARMD

revision surgery have likely improved over time. Possible explanations for this improvement were lowered threshold for performing revision surgery, improved patient surveillance, more experienced surgeons and longer time interval from primary operation to revision. The lowest rates of complications and re-revisions were reported in the largest studies with experienced, high-volume surgeons.

## 2.5 Histopathology of ARMD

### 2.5.1 Histology and organization of healthy synovium

The synovium forms an enclosed environment for the joint. It consists of a synovial membrane that attaches to the bone surfaces. The functions of the synovium include lubrication of the joint and defense against pathogens. The synovial membrane can further be divided into intimal and subintimal layers. The intimal layer, also termed the synovial lining, is the innermost layer. Intimal cells are known as synoviocytes and can be divided into two major types: A and B. Type A synoviocytes are macrophage-like and form an immunological barrier. Type B synoviocytes resemble fibroblasts and produce important components of synovial fluid. Under the intima lays the subintima, which consists of the extracellular matrix and a few cells. Occasional inflammatory cells, such as macrophages and lymphocytes, are also observed. (Smith 2011). A photomicrograph of a histological section of hip joint synovium is shown in Figure 6.
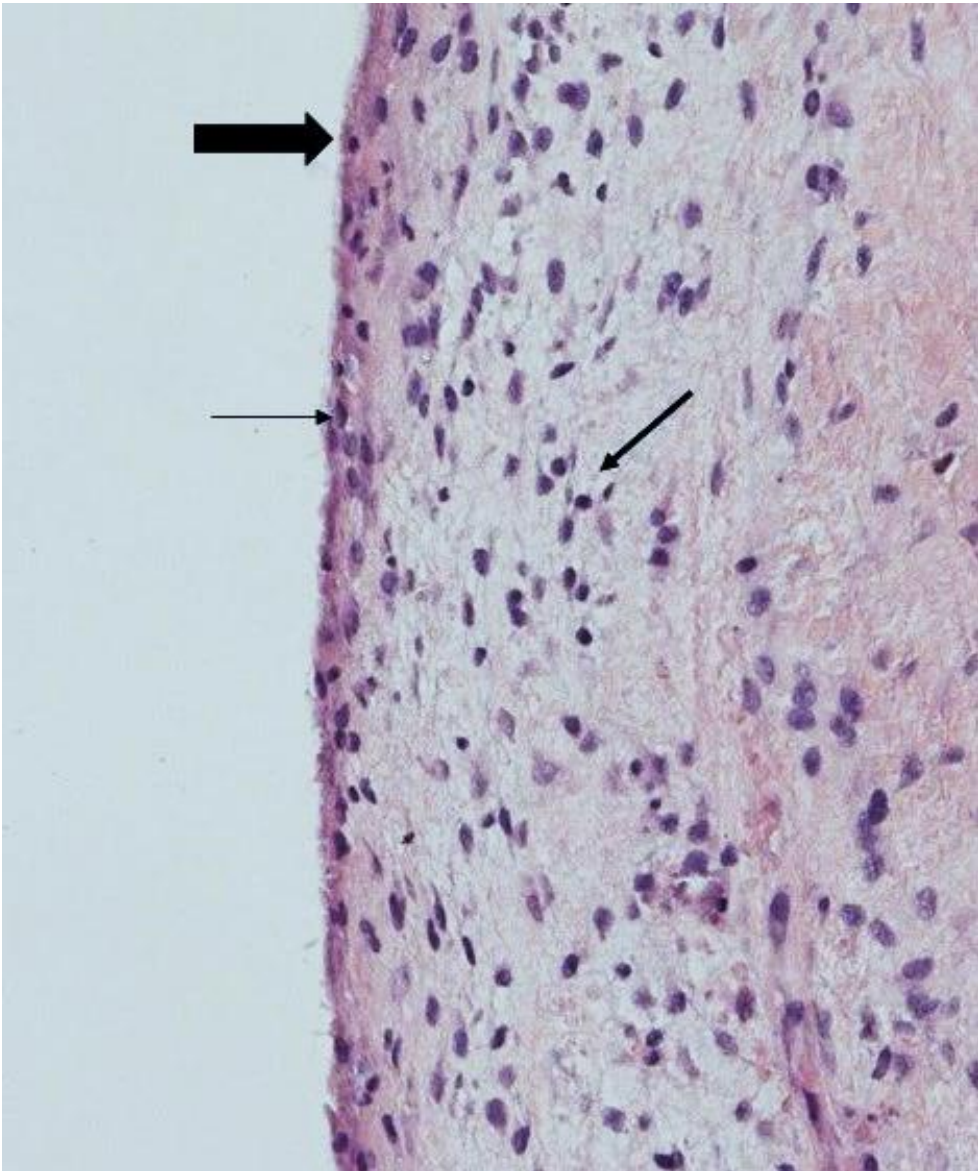
**Figure 6.** A photomicrograph of a H&E stained section of healthy hip joint synovium. The thick arrow points to intact synovial lining. The thin arrow points to synoviocytes. The intermediate arrow shows subintimal macrophages and fibroblasts. Photomicrograph captured with Nikon Eclipse 50i fitted with 20x objective (total magnification 200x).

## 2.5.2 Overview of the histopathology of ARMD

Several studies have been conducted regarding the histopathology of ARMD. In these studies, samples, such as synovial capsule and pseudotumor tissue, have been obtained from the periprosthetic tissues and graded using light microscopy. Some studies have used flow cytometry to determine the number of inflammatory cells present.

Findings observed in periprosthetic tissue samples obtained from patients with ARMD include necrosis, inflammatory cell infiltrates consisting of variable amounts of neutrophils, macrophages, plasma cells, T- and B-lymphocytes, germinal centers, sarcoid-like granulomas and vascular damage (Davies et al. 2005, Willert et al. 2005, Campbell et al. 2010, Natu et al. 2012, Grammatopoulos et al. 2013). Most of the inflammatory cells are either macrophages or lymphocytes. Necrosis may be coagulative or fibrinoid (Krenn et al. 2014). In addition, there appears to be distinct patterns of histopathological findings. At least three different pathological responses have been suggested: 1. Foreign-body macrophage response, 2. Cytotoxic response and 3. ALVAL response (Davies et al. 2005, Willert et al. 2005, Mahendra et al. 2009, Campbell et al. 2010, Natu et al. 2012, Berstock et al. 2014). It should be noted, however, that these responses may overlap considerably and may present simultaneously (Berstock et al. 2014, Ricciardi et al. 2016). Several different scoring systems for tissue samples have been used of which the ALVAL score is likely the most popular (Campbell et al. 2010, 2018b). ALVAL scoring and Natu scoring are explained in more detail in the Methods section, chapter 4.2.4.

## 2.5.3 Foreign-body macrophage response

It is well recognized that the generation of polyethylene particles in patients with MoP implants induces an inflammatory macrophage response which can, over time, lead to osteolysis and aseptic loosening of the implant. The degree of osteolysis is related to the high volume of polyethylene wear debris produced. (Harris 1994). Although MoM implants are substantially lower wearing, metal particles also appear to evoke a macrophage response in the synovial tissue in some patients, although milder than that seen with MoP implants (Doorn et al. 1996, Willert et al. 1996, 2005). Macrophages have three major roles in tissue inflammation: phagocytosis of foreign particles or not viable tissues, antigen presentation and modulation of the immune response via numerous cytokines and

growth factors (Fujiwara and Kobayashi 2005). Histopathologically, this type of response in the synovium is characterized by the presence of superficially located macrophages with or without foreign-body giant cells and granulomas (Figure 7) and preservation of the underlying tissue architecture (Campbell et al. 2010, Berstock et al. 2014). Fine metal particles or "metal dust" are often seen in the cytoplasm of macrophages as a sign of phagocytosis (Nawabi et al. 2014) (Figure 8). Immunologically, metal particles are recognized by macrophages as foreign material, which activates numerous signaling pathways to gather more macrophages to the site to help clear the metal debris. Diffuse lymphocytes may also be present in minor quantities. This type of macrophage-mediated response is termed innate or non-specific immune response. (Athanasou 2016). For a microscope image of the characteristic histological features, please see Figure 11, page 100.



**Figure 7.** Photomicrograph of an H&E stained section of synovial tissue removed from patient revised for ARMD. The two-headed arrow shows the whole granuloma with metallic debris encapsulated inside. One-headed arrow points to a multinucleated giant cell. Photomicrograph captured with Nikon Eclipse 50i fitted with 20x objective (total magnification 200x).

**Figure 8.** Photomicrograph of an H&E stained section of synovial tissue removed from patient revised for ARMD. The arrow points to a macrophage with phagocytized fine metal debris inside. Photomicrograph captured with Nikon Eclipse 50i fitted with 40x objective (total magnification 400x).

## 2.5.4 Cytotoxic response

In some patients, the histological picture of ARMD comprises substantial tissue necrosis in addition to macrophage infiltration. It has been suggested that in these patients the underlying cause of necrosis is the direct cytotoxic effect of cobalt-chromium particles and ions (Mahendra et al. 2009). The presumed mechanism is the following: metal particles are phagocytized by periprosthetic cells and contained in phagolysosomes, the acidic environment of lysosomes leads to degradation of the metal particles into ions, which then escape the lysosome and lead to apoptosis and necrosis of the affected cells (Salvati et al. 1993, Xia et al. 2011). In keeping with this hypothesis are in vitro studies which show that cobalt and chromium ions can cause dose-dependent necrosis and apoptosis in macrophages (Catelas et al. 2001, 2005, Kwon et al. 2009). Several authors have

suggested that the necrosis induced by metal ions leads to a cycle in which macrophages are first recruited to clear the cellular and metallic debris but end up undergoing cell death themselves, which then leads to more macrophages being recruited and a worsening of the situation (Salvati et al. 1993, Mahendra et al. 2009, Grammatopoulos et al. 2013). Further, there is evidence that the release of metal ions and particles from dead macrophages leads to the cell death of adjacent fibroblasts as well (Xia et al. 2011). It is not understood why some patients develop necrosis in addition to macrophage inflammation, while some do not (Eltit et al. 2019). In conclusion, the cytotoxic response is likely a combination of the direct cytotoxic effect of metal debris and the activation of the innate immune response which leads to the histopathological findings of tissue necrosis and heavy macrophage infiltration. For the characteristic features of this response, please see Figure 12, page 101.

## 2.5.5    ALVAL response

Willert et al. and Davies et al. were the first to report that in some patients with early onset of pain the capsular tissues display prominent lymphocytic infiltration, necrosis, fibrin exudation, vascular wall changes, occasional plasma cells and variable amounts of macrophages (Davies et al. 2005, Willert et al. 2005). Lymphocytes were present diffusely in the superficial layer and as perivascular cuffs in the intermediate layer of the capsule. The authors suggested the presence of lymphocyte-dominated type IV hypersensitivity response to metal debris and described it as Aseptic Lymphocyte-dominated Vasculitis-Associated Lesion (ALVAL). Perivascular lymphocytic cuffing of the capillaries, swelling of the vascular walls and necrosis are characteristics for vasculitis, and thus the findings were described as vasculitis-associated lesion. Davies et al. noted, however, that it is unclear whether these findings represent active vasculitis or a novel form of immunological response with unknown consequences.

Since the pioneering work by Davies et al. and Willert et al., numerous studies have described mostly similar findings (Witzleb et al. 2007, Huber et al. 2008, Pandit et al. 2008b, Mahendra et al. 2009, Campbell et al. 2010, Natu et al. 2012, Grammatopoulos et al. 2013, Berstock et al. 2014, Langton et al. 2016, Ricciardi et al. 2016). The role of vascular wall changes in the development of tissue necrosis has been questioned (Mahendra et al. 2009, Natu et al. 2012). Natu et al. suggested that necrosis is likely due to pronounced lymphocytic inflammation but may also

be due to vascular wall changes leading to local ischemia and necrosis. The authors stated that it is still not understood whether the vascular wall changes are a consequence of lymphocytes transiting through the wall or true vasculitis. In a study by Langton et al., it was shown that the thickness of lymphocytic cuffing correlated with the degree of necrosis (Langton et al. 2011b). Further, T-killer lymphocytes, which can cause necrosis, have been found in tissues with suggested type IV response and may be related to the common finding of tissue necrosis in ALVAL response (Hasegawa et al. 2016). In some patients with lymphocytic tissue responses, germinal centers containing B- and T-lymphocytes have also been observed (Natu et al. 2012, Langton et al. 2013b, Mittal et al. 2013). A microscope image of germinal center is shown in Figure 9. Natu et al. suggested that these are the end stage of lymphocytic inflammatory response. However, Mittal et al. concluded that these germinal centers, or tertiary lymphoid organs, form their own distinct pathological subset of ARMD. The role of these germinal centers is not fully understood (Hasegawa et al. 2016). Altogether, studies have supported the hypothesis that an adaptive immune response, leading to lymphocyte-dominated inflammation and subsequent necrosis, is the cause of failure in some patients. For a microscope image of characteristic histological features, please see Figure 13, page 102.

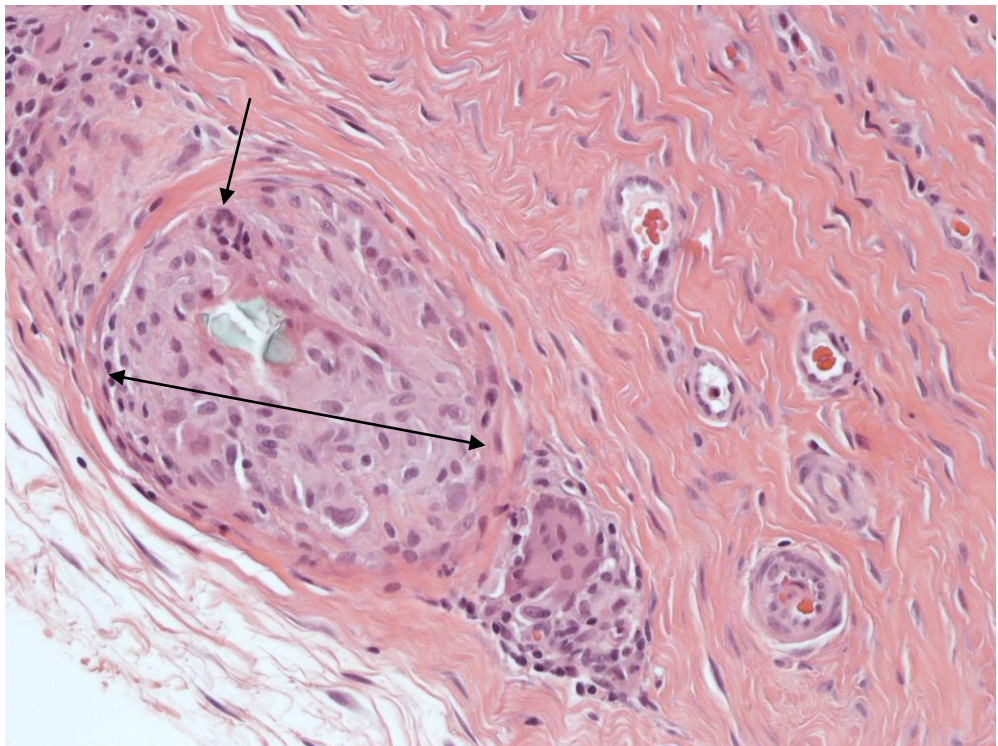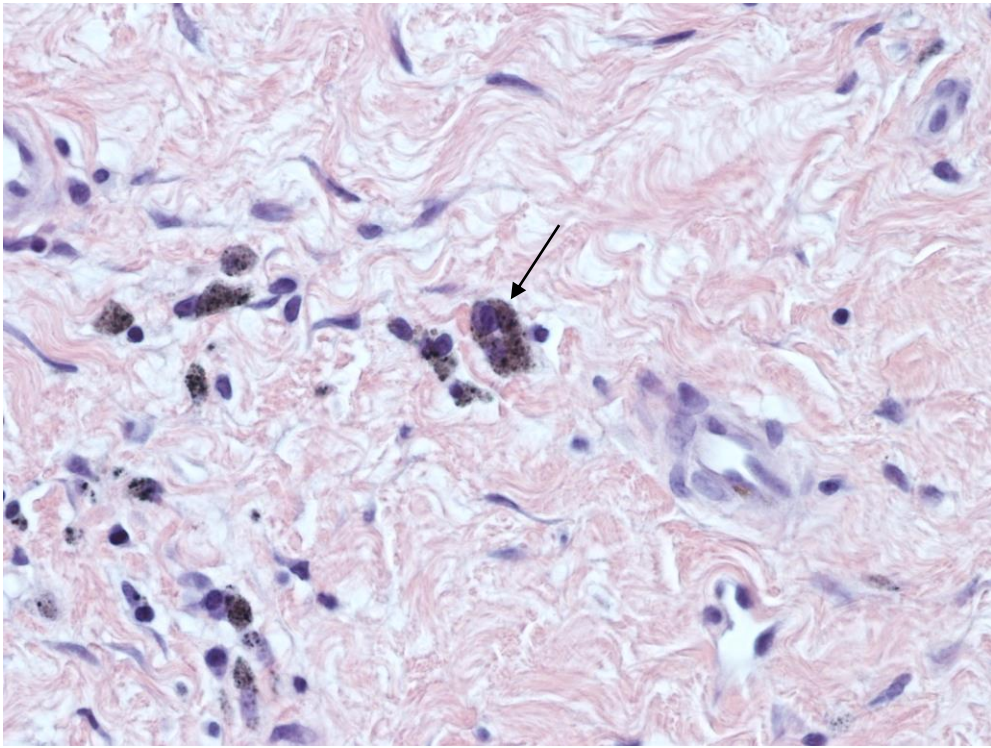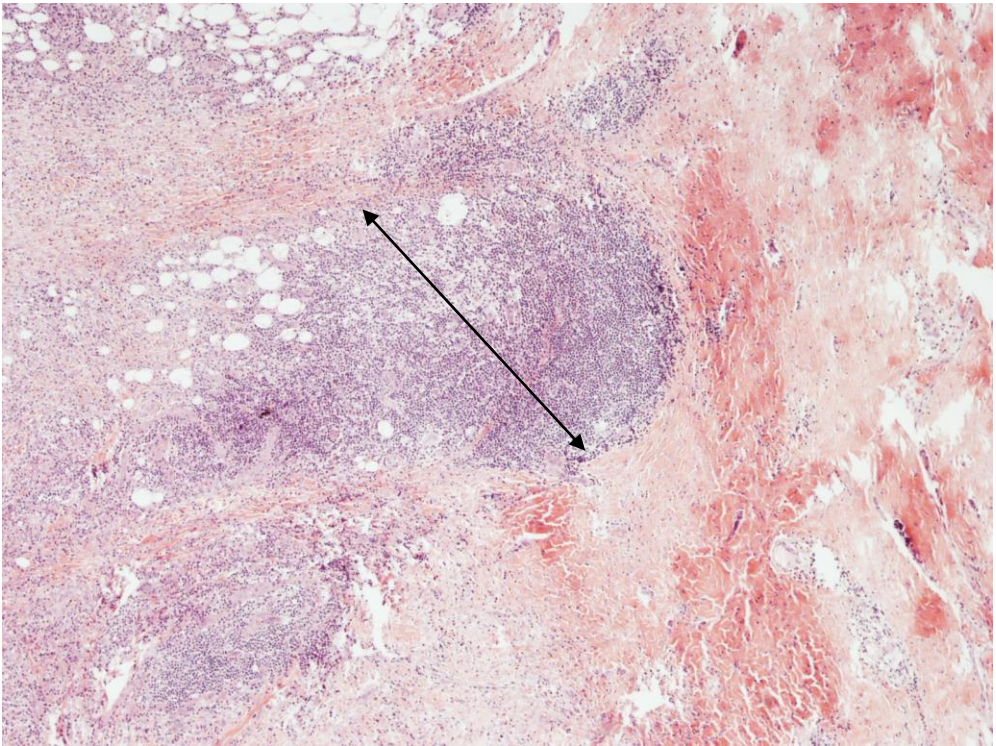**Figure 9.** Photomicrograph of an H&E stained section of synovial tissue removed from patient revised for ARMD. The arrow shows a germinal center. Photomicrograph captured with Nikon Eclipse 50i fitted with 4x objective (total magnification 40x).

## 2.5.6    Histopathology of pseudotumors

The term pseudotumor is often used interchangeably with the term ARMD. However, by definition, pseudotumors are cystic or solid masses that connect with the joint space (Pandit et al. 2008a). It is not understood why pseudotumors form in only a subset of patients who develop ARMD. Pseudotumors can be associated with all of the responses described above – Foreign-body, cytotoxic and ALVAL response (Mahendra et al. 2009, Campbell et al. 2010, Grammatopoulos et al. 2017a). Further, pseudotumors do not seem to be specific to MoM implants as they are also observed in patients with other types of bearing couples (Carli et al. 2011). In patients with MoP hip implants, corrosion debris from the trunnion most likely causes the formation of pseudotumors (Cooper et al. 2012, 2013). However, pseudotumors may also form in response to polyethylene wear (Murgatroyd 2012). The histology of these lesions appears to be solely granulomatous macrophage

inflammation and no lymphocytes have been observed (Carli et al. 2011). This is in contrast to the common finding of heavy lymphocytic infiltration and necrosis in MoM pseudotumors (Mahendra et al. 2009, Campbell et al. 2010).

## 2.6    Etiopathogenesis of ARMD

As reviewed above, three main types of tissue responses have been suggested in ARMD, and these may further overlap. Periprosthetic tissues show inflammatory changes, such as swelling, presence of macrophages, T-lymphocytes and necrosis, in different proportions (Grammatopoulos et al. 2013, Berstock et al. 2014). Other findings include occasional B-lymphocytes, plasma cells, neutrophils, granulomas and germinal centers (Natu et al. 2012, Grammatopoulos et al. 2013, Hasegawa et al. 2016). These different microscopic, tissue-level findings reveal important indirect information about the underlying pathological processes. It is for this reason that the histopathological analysis of tissues related to failed MoM hip replacements is important. (Athanasou 2016). Another, less commonly used method for evaluating the cellular response is flow-cytometry. This method uses specific cellular antigens to detect the exact amounts of different inflammatory cells and their subgroups present in a tissue sample (Brown and Wittwer 2000).

The most commonly scored histological features are the presence and extent of macrophages, lymphocytes and necrosis (Campbell et al. 2018b). Macrophages are considered a key component of the innate, non-specific response to foreign material. Lymphocytes, on the other hand, form the cellular basis of the adaptive immune response, which is antigen-specific and has an immunological memory. (Athanasou 2016). Necrosis may be the result of the direct cytotoxic effects of metal wear but may also be due to lymphocytic inflammation (Mahendra et al. 2009, Campbell et al. 2010, Langton et al. 2011b). Innate and adaptive immunological responses are not exclusive. The adaptive response may then be provoked if foreign material (metal debris) is presented by antigen-presenting cells and recognized as a specific antigen by sensitized lymphocytes. Macrophages help maintain inflammation by secreting cytokines that gather and affect lymphocytes in many ways (Athanasou 2016). In regard to the ALVAL response, it has been suggested that the underlying mechanism is an adaptive, cell-mediated type IV response that leads to the accumulation of diffuse and perivascular lymphocytes (mainly T-lymphocytes), inflammation and necrosis of the periprosthetic tissues (Davies et al. 2005, Willert et al. 2005). It has been postulated that metal ions may

form complexes with host-proteins which would change their conjugation, be recognized as foreign antigens and lead to activation of the adaptive immunological cascade (Athanasou 2016, Eltit et al. 2019).

Histopathological findings may be further combined with clinical and retrieval information, such as implant wear, gender and time from primary operation to revision surgery, to better understand the underlying pathogeneses and their variations in individuals (Campbell et al. 2018b). Since ARMD is considered a consequence of wear debris, many studies have focused on the associations between metal wear burden and histopathological findings (Table 5) (Langton et al. 2010, Takamura et al. 2014). Some studies have directly measured wear from retrieved implants, and some have used indirect methods, such as blood/synovial metal ion concentrations or periprosthetic tissue metal concentration (Table 5). As becomes evident from the table, the results of these studies have been discrepant.

In a central study in 2010, Campbell et al. showed that the ALVAL response was associated with low implant wear. Conversely, a macrophage-dominant response was seen with high implant wear. This formed a basis for a hypothesis that the ALVAL response is due to hypersensitivity to metal debris and does not therefore require abnormal amounts of metal debris to be provoked (Campbell et al. 2010). This hypothesis has been supported by the majority of the research; however, opposing results have also been published (Table 5). On a group-level, associations between metal wear burden and type of tissue response have mostly been weak. In some studies in which group-level associations were not found between metal wear burden and lymphocytes, the authors noted that a subset of patients presented features of ALVAL responses and had low wear, supporting the original hypothesis (Grammatopoulos et al. 2013, 2017a).

Campbell et al. suggested that high amounts of metal wear lead to a foreign-body, macrophage-dominated innate response (Campbell et al. 2010). Many of the studies have supported this finding (Table 5). Grammatopoulos et al. further observed that high wear was associated with both macrophages and necrosis (Grammatopoulos et al. 2013). They proposed a cycle where metal particles cause direct cytotoxic effects to cells which undergo cell-death. This leads to the accumulation of macrophages to clear the cell debris and results in the macrophages also facing cell-death. This subsequently leads to the recruitment of more macrophages. The direct cytotoxic effects of metal particles have also been suggested by another study (Mahendra et al. 2009). However, it remains unclear whether the foreign-body macrophage-response without necrosis is a separate entity.

Other etiopathological factors, such as particle size and origin, patient susceptibility and gender, have also been suggested. Metal debris from the taper-interface is potentially more immunogenic and cytotoxic than debris from the bearing surfaces (Langton et al. 2013a, Xia et al. 2017). Xia et al. reported that taper debris led to more severe necrosis and lymphocytic infiltration than bearing wear despite smaller amounts of metals being present in the tissues. Another often proposed factor is patient susceptibility. It has been suggested that individual reactivity to metal debris is variable (Grammatopoulos et al. 2013). The existence of individual patient susceptibility is widely accepted (Campbell et al. 2014). There is, however, no direct evidence to support this belief. Studies have been conducted regarding clinical testing for metal hypersensitivity, but no clinically useful tests have been found (Teo and Schalock 2016). Women have been observed to be at higher risk for failure at similar levels of blood metal ions compared to men (Langton et al. 2013b). Other researchers have published similar findings, as discussed earlier in Chapter 2.4.3. Langton et al. also noted that the ALVAL response was overrepresented in women. They suggested that when compared to men, women may be more prone to mount adaptive immune responses leading to ARMD.

The pathogenesis of ARMD is still poorly understood. Several different mechanisms have been proposed, but many of the studies regarding etiopathogenesis have been in disagreement. Furthermore, the observed associations have mostly been relatively weak. In conclusion, there appears to be several different pathological entities which most often lead to ARMD in the presence of a high wearing implant. However, ARMD is also encountered in patients with low wearing implants, and the presence of an adaptive, lymphocytic ALVAL response in these patients is supported by the body of evidence. These responses often overlap and other, still unrecognized, responses may exist (Grammatopoulos et al. 2013, Berstock et al. 2014, Ricciardi et al. 2016)

**Table 5.** The reported associations between metal measurements and tissue responses in the published literature.

| Study | Metal measurement | Tissue characterization | N | Association between metal measurement & tissue response | | |
|---|---|---|---|---|---|---|
| | | | | Macrophages | Lymphocytes | Necrosis |
| Campbell et al. 2010 | Linear wear | Histology | 32 | High wear | Low wear | Not associated |
| Langton et al. 2011a | Volumetric wear | Histology | 25 | Not associated | Not associated | Not associated |
| Ebramzadeh et al. 2011 | Linear wear | Histology | 353 | - | Low wear | - |
| Grammatopoulos et al. 2013 | Volumetric wear | Histology | 56 | High wear | High wear | High wear |
| Ebramzadeh et al. 2014 | Linear wear | Histology | 119 | High wear | Low wear | Not associated |
| Nawabi et al. 2014 | Volumetric wear | Histology | 94 | - | Low wear | - |
| Campbell et al. 2018a | Volumetric wear | Histology | 165 | High wear | Not associated | Not associated |
| Lohmann et al. 2013 | Synovial tissue concentration | Histology | 28 | Low metal load | High metal load | - |
| -"- | Serum ion concentrations | -"- | 27 | Not associated | Not associated | - |
| Reito et al. 2015b | Synovial fluid concentrations | Histology | 163 | Not associated | High metal load | High metal load |
| Paukkeri et al. 2016 | Whole blood concentrations | Flow cytometry | 16 | High metal load | Low metal load | - |
| Grammatopoulos et al. 2017a | Whole blood concentrations | Histology | 38 | Not associated | Not associated | Not associated |

# 3   AIMS

The purpose of this dissertation was to investigate the histopathological characteristics of ARMD, implant wear, clinical markers of implant wear, periprosthetic tissue metal ion levels, and to investigate the relationships between these characteristics in order to better understand the etiopathogenesis of ARMD.

The specific aims of the studies were to investigate:

Study I:        Bearing wear and its association with histopathological findings in patients with failed ASR MoM hip resurfacings

Study II:       The association between periprosthetic tissue metal ion levels, whole blood and synovial fluid metal ion levels and histopathological findings in patients with failed MoM hip replacements

Study III:      The histopathological patterns and possible subgroups in ARMD

Study IV:       The role of host-specific factors in the pathogenesis of ARMD

# 4 PATIENTS AND METHODS

## 4.1 Patients

Between November 2000 and February 2012, a total of 2 868 primary MoM hip replacements (LD THA and hip resurfacings) were implanted in 2 398 patients at our institution. Of these, 1 036 were ASR MoM hip replacements (Depuy Orthopaedics, Warsaw, IN, USA) that were implanted in 887 patients between March 2004 and December 2009. For the purposes of the studies included in this dissertation, we retrospectively identified patients who had undergone revision surgery of a MoM hip replacement either unilaterally or bilaterally. We did not perform power calculations since all of our studies were retrospective in nature. All patients with sufficient data were included. A summary of the studies is presented in Table 6.

For study I, we identified those patients with ASR hip resurfacings that had been revised at our institution between the manufacturer's recall of the ASR hip replacements in August 2010 and the end of our recruitment period in January 2016. During this period, a total of 114 ASR hip resurfacing devices in 107 patients were revised at our institution. All consecutively revised patients who gave informed consent and fulfilled the following criteria were included in our study: 1) Revision was due to ARMD, 2) Retrieved components were available for bearing wear analysis and 3) A periprosthetic tissue sample was available for histopathologic analysis. After exclusion, 85 hips in 78 patients were included in our study. Twenty-one of these patients were referred to our institution from central hospitals from other hospital districts and 57 patients had had their index operation (primary arthroplasty) and follow-up at our institution. Of the 85 hips included in the study, 56 were explanted from female patients and 29 from male patients. Mean age at the time of the revision surgery was 57.3 years (SD 10.3 years). Mean follow-up time between index operation and revision surgery was 5.4 years (SD 1.8 years). Surgery was performed by or under the direct supervision of 14 senior orthopaedic surgeons.

For study II, we recruited a pilot patient in June 2013 followed by the recruitment of consecutive patients between February 2014 and August 2016. In total, 134 hips with variable MoM implants were revised for ARMD at our institution during the recruitment period. Of these, two hips were not included due to infection, two hips were not included due to inadequate tissue samples and 23 hips were not included as they were operated on by surgeons who did not participate in the recruitment and sample collection. Thus, in total, 107 hips were included. Of these, 87 were THA and 20 were hip resurfacings. Whole blood samples were available for 106 patients and synovial fluid samples for 90 patients. In addition to the patients undergoing revision surgery, two further patients who had undergone primary hip arthroplasty and whose tissue samples had been retrieved from osteoarthritic synovium were recruited as controls for tissue metal analysis. The revised implants for the study are presented in more detail in Table 7.

For study III, we identified patients who had been implanted with an ASR MoM hip replacement (either THA or hip resurfacing) at our institution and had undergone revision surgery between the recall of the ASR hip replacements in August 2010 and the end of our recruitment period in January 2015. During this time, a total of 334 hips in 301 patients were revised. In 296 of the 334 revision surgeries, a tissue sample was retrieved for histopathological analysis, and these patients were included in the study. The vast majority of these patients were revised for ARMD.

For study IV, we included all patients with an ASR hip replacement (either THA or hip resurfacing) who had undergone bilateral revision surgery at our institution. By the end of our recruitment period in September 2016, 316 patients had been revised. Of these, 33 patients had undergone bilateral revision. Four of these patients were excluded due to missing tissue samples. Thus, 29 patients were included in our study (58 hips). All patients had the same head-cup-combination on both sides: five patients had bilateral ASR hip resurfacing and 24 patients had ASR XL stemmed THA bilaterally. Simultaneous bilateral hip revision was performed for two patients, and bilateral revision surgeries were performed for the remaining 27 patients sequentially.

**Table 6.**    Summary of study designs, patient demographics and retrieved hip replacements.

| Study | I | II | III | IV |
|---|---|---|---|---|
| **Study design** | Retrospective cohort | Retrospective cohort | Retrospective cohort | Retrospective cohort |
| **Patients** | 78 | 99 | - | 29 |
| **Hips** | 85 | 107 | 296 | 58 |
| **Females (percentage of all patients)** | 56 (66%) | 42 (42%) | - | 13 (45%) |
| **Mean age at primary operation (SD)** | 57.3 (10.3) | 66.8 (7.5) | - | *First hip 61.7 (8.3) *Second hip 63.1 (8.5) |
| **Mean follow-up time (SD)** | 5.4 (1.8) | 7.1 (2.5) | - | First hip 4.5 (1.29) Second hip 5.8 (1.8) |
| **Retrieved hip replacements** | ASR resurfacing | See Table 7 | ASR resurfacing and THA | ASR resurfacing and THA |

*Mean age at revision operation

**Table 7.**        Retrieved hip replacements in Study II.

**Retrieved total hip replacements**

| Femoral component | Acetabular component | Quantity |
|---|---|---|
| DePuy Summit | DePuy ASR | 32 |
| DePuy Summit | DePuy Pinnacle | 10 |
| Biomet Bimetric | Biomet M2A38 | 10 |
| Biomet Bimetric | Biomet ReCap | 5 |
| DePuy Corail | DePuy ASR | 4 |
| Smith-Nephew Synergy | Smith-Nephew R3 | 4 |
| Zimmer ZMR | Zimmer Durom | 2 |
| Zimmer M/L Taper | Zimmer Durom | 2 |
| Wright Medical Profemur | Wright Medical Conserve Plus | 2 |
| Other | Other | 16 |
| | | Total = 87 |

**Retrieved hip resurfacings**

| | Quantity |
|---|---|
| DePuy ASR | 9 |
| Smith-Nephew BHR | 6 |
| Zimmer Durom | 2 |
| Biomet ReCap | 2 |
| Smith-Nephew BHR – TM Revision shell | 1 |
| | Total = 20 |

## 4.2 Methods

### 4.2.1 Follow-up of MoM patients at our institution

After the recall of the DePuy ASR hip arthroplasties and the Medicines and Healthcare products Regulatory Agency (MHRA) medical device alert regarding MoM hip arthroplasties, a systematic screening program was launched at our institution (DePuy Orthopaedics 2010, MHRA 2010). All patients with MoM hip arthroplasty were included in the program. Patients were given an Oxford Hip Score questionnaire to assess their symptoms and hip function (0-48p, 48p best possible), examined physically (including the Harris Hip Score) and whole blood (WB) chromium and cobalt ion levels were measured (Harris 1969, Dawson et al. 1996). Hip and pelvic radiographs were taken before each visit. In addition, all ASR patients were referred for Metal Artifact Reduction Sequence MRI (MARS-MRI), unless there were contraindications, in which case patients were referred for ultrasound imaging of the hips. Patients with hip replacements other than ASR were referred for MARS-MRI or ultrasound imaging if they had symptoms or elevated WB metal ion levels. Findings were classified using a previously published Imperial pseudotumor classification in which pseudotumors are graded 1, 2A, 2B or 3, depending on their wall-thickness, contents and shape (Table 8) (Hart et al. 2012d).

**Table 8.**  Imperial grading of pseudotumors using MARS-MRI. Adopted from Hart et al. 2012d.

| Pseudotumor type | Wall | Contents | Shape |
|---|---|---|---|
| 1 | Thin-walled | Fluid-like: hypointense on T1, hyperintense on T2 | Flat, with walls mainly in apposition |
| 2A | Thick-walled or irregular | Fluid-like: hypointense on T1, hyperintense on T2 | Not flat, with >50% of the walls not in apposition |
| 2B | Thick-walled or irregular | Atypical fluid: hyperintense on T1, variable on T2 | Any shape |
| 3 | Solid throughout | Mixed signal | Any shape |

## 4.2.2    Indications for revision surgery

Revision surgery was considered if 1) a clear pseudotumor (class 2A, 2B or 3) (Hart et al. 2012d) was observed on cross-sectional imaging regardless of symptoms or WB metal ion levels; or 2) the patient had elevated WB metal ion levels and hip symptoms despite normal findings in cross-sectional imaging; or 3) the patient had a continuously symptomatic hip or progressive symptoms regardless of imaging findings or metal ion levels; or 4) the patient had progressively increasing blood metal ion levels even without symptoms or findings in cross-sectional imaging (Reito et al. 2013). Symptoms included hip pain, discomfort, sense of instability, and/or impaired function of the hip and sounds from the hip (clacking, squeaking). WB metal ion levels were regarded as being elevated if either chromium or cobalt exceeded 5 ppb (Hart et al. 2011b).

## 4.2.3    Definition of ARMD in this thesis

Failure was classified as being due to ARMD on the basis of the following criteria (Reito et al. 2013, Lainiala et al. 2015): 1) there was presence of metallosis or macroscopic synovitis in the joint; and/or 2) a pseudotumor was found during revision; and/or 3) a moderate to high number of perivascular lymphocytes along with tissue necrosis and/or fibrin deposition was seen in the histopathologic sample; and 4) perioperatively, there was no evidence of component loosening or periprosthetic fracture. In addition, infection was ruled out by obtaining multiple (at least five) bacterial cultures during revision surgery.

## 4.2.4    Histopathological analysis of periprosthetic tissue (Studies I, II, III and IV)

During every MoM hip revision at our institution, samples of the inflamed synovia and/or pseudotumor capsule are obtained for histopathological analysis. One to five samples are taken, depending on the surgeon and the availability of excess periprosthetic tissue. The samples are obtained from the most metallotic, inflamed synovial capsule and/or pseudotumor capsule. The results of these analyses were used in all the studies of this dissertation (I, II, III and IV). For analysis, each obtained tissue sample was formalin fixed and embedded in paraffin. Several 10 μm microtome sections were made and stained with standard hematoxylin and eosin

staining. The sections were examined histologically under transmitted light with a Nikon Eclipse 50i microscope (Nikon Corporation, Shinagawa, Tokyo, Japan). The sections were graded by a senior musculoskeletal pathologist (Jyrki Parkkinen) with more than 10 years' experience in the field, using scoring principles adopted from the study by Natu et al. (Natu et al. 2012). If the severity of findings between different sections of the same tissue sample were inconsistent, averages were determined. In study I, a more concise ALVAL scoring system was used in addition to Natu grading (Campbell et al. 2010). The pathologist was blinded from clinical patient characteristics.

The Natu grading comprised the following parameters: 1) macrophage sheet thickness, 2) perivascular lymphocyte cuff thickness, 3) degree of necrosis, 4) presence of plasma cells, 5) presence of diffuse lymphocytic infiltrate, 6) presence of germinal centers and 7) presence of granulomas. The thickness of macrophage sheets was calculated using a graticule and graded 0–3 (absent, < 1 mm, 1–2 mm, > 2 mm). Lymphocyte cuff thickness was also calculated using a graticule. An average of five measurements was taken and graded as 0–3 (absent, 0.25 mm, 0.25–0.75 mm, > 0.75 mm). The extent of overall tissue necrosis in a sample was graded based on the surface necrosis typing according to Davies et al. (Davies et al. 2005). Type 1 surface contains intact synovial epithelium. Type 2 surface shows loss of synovial epithelial cells without fibrin deposition. In type 3 surface, there is fibrin deposition, and in type 4 surface there is extensive necrosis and loss of architecture. The extent of type 4 surface necrosis was used to grade the overall tissue necrosis in a given sample, as described by Natu et al. In grade 4 necrosis, more than 75% of the tissue sample showed type 4 surface necrosis. In grade 3 necrosis, between 25 and 75% showed type 4 surface necrosis. In grade 2 necrosis, either less than 25% of the tissue showed type 4 surface necrosis or the tissue showed type 3 surface. In grade 1 necrosis, the sample consisted of type 2 surface.

In ALVAL grading, a total score of 0-10 is given based on three subscores: synovial lining (0-3p), tissue organization (0-3p) and inflammatory infiltrate (0-4p). Both synovial lining and tissue organization reflect the degree of necrosis and higher scores mean a higher degree of necrosis. Inflammatory infiltrate score reflects the predominant inflammatory cell type on a spectrum: 0 points means minimal infiltrates, 1 point means predominantly macrophages, 2 points means both macrophages and diffuse/perivascular lymphocytes, 3 points means mostly lymphocytes in aggregates and some macrophages and 4 points means large lymphocyte aggregates and little to no macrophages. Total scores of 0-4 points are considered low, 5-8 moderate and 9-10 high. The subscores graded are thought to

reflect the features of a prominent ALVAL response. The authors suggested that high scores distinguish hypersensitivity-related ALVAL responses from those responses related to high wear (low scores). (Campbell et al. 2010).

## 4.2.5    Metal analysis of the periprosthetic tissue (Study II)

In study II, metal concentrations were analyzed from the obtained periprosthetic tissue samples in addition to histopathological grading. For metal content analysis, a subsample (approx. 0.3 g) was cut from the tissue sample, weighed and transferred into a Teflon vessel. Samples were first decomposed with 5 ml Suprapur $HNO_3$ (Merck) by microwave digestion technique using a CEM MDS-2000 Microwave System (CEM corporation, Matthews, NC, USA) and then diluted to 10 ml with Milli Q-water. The digests were analyzed for Al, Cr, Co, Ti, Mo and V with an Inductively Coupled Plasma Optical Emission Spectrometer. A Thermo Electron iCAP 6600 Duo View equipped with Cetac ASX-520Hs and autosampler was used (Thermo Fisher Scientific, Waltham, MA, USA). Detection limits for Al, Cr, Co, Ti, Mo and V were 9.0, 0.2, 0.2, 3.0, 0.2 and 3.0 µg/g, respectively. NIST SRM 1576b (Bovine liver) was used as a certified reference material to ensure the performance of the analytical procedure for tissue samples.

## 4.2.6    Whole blood metal ion measurement (Studies I, II and III)

In studies I, II and III, we utilized the metal ion measurements conducted in the follow-up of patients for investigational purposes. All patients underwent WB analysis of Co/Cr following sampling from the antecubital vein using a 21-gauge needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) and trace-element blood tubes containing sodium ethylenediaminetetraacetic acid (EDTA). Standard operating procedures were established at the Finnish Institute for Occupational Health for Co and Cr measurement using dynamic reaction cell inductively coupled plasma (quadripole) mass spectrometry (Agilent 7500 cx, Agilent Technologies, Santa Clara, CA, USA). The laboratory technicians were blinded to all clinical outcomes. The samples were preserved at +6 °C to +8 °C prior to analysis.

### 4.2.7 Synovial fluid metal ion measurement (Study II and III)

In studies II and III, synovial fluid (SF) analysis was performed. Since October 2011, our MoM hip revision protocol has involved perioperative SF aspiration, which is always taken before opening the deep fascia using a standard 18- to 20-gauge needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, New Jersey) and trace element tubes containing sodium EDTA. Similar procedures as those described above for WB were used for SF metal ion concentration measurement.

### 4.2.8 Bearing wear analysis (Studies I and IV)

In studies I and IV, the volume of material loss from the cup and head bearing surfaces was measured at the London Implant Retrieval Center (LIRC) using a Zeiss Prismo (Carl Zeiss Ltd., Rugby, UK) coordinate measuring machine (CMM). A total of 400 polar scan lines on each surface were defined and up to 30 000 data points captured using a 2 mm ruby stylus; protocols for this method have been previously published (Bills et al. 2012). An iterative least square fitting method was used to analyze the raw data captured by the CMM and the unworn geometry of the bearing surface was used to map regions of material loss from which the total volumetric loss was calculated for each component. Total wear volume was calculated by combining head and cup wear volumes for each patient. Volumetric wear rate ($mm^3$/year) was further calculated by dividing total wear volume in cubic millimeters by implantation time in years.

### 4.2.9 Statistical analysis

In all but study III, statistical analyses were conducted using SPPS software (IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). In study III, analyses were performed using R Software, version 3 (R Foundation, Vienna, Austria). In all studies, mean values with standard deviation (SD) were reported for normally distributed variables and medians with range and/or interquartile range (IQR) for variables with skewed distribution. Differences between non-normally distributed variables were compared using Mann-Whitney U-test. In all studies, p-values of $<0.05$ were considered statistically significant.

In studies I and II, Spearman rank correlation was used to study the associations between different variables due to non-normal distribution of these variables. When analyzing the correlation between WB metal ion concentrations and other factors, we only included patients with unilateral hip arthroplasties to avoid the confounding effect of metal ions being released to the blood from the other implant.

In study III, two different cluster-based segmentation methods were used to detect the underlying latent groupings of cases. Latent class analysis (LCA) and cluster analysis with hierarchical approach (HCA) were used (Beckstead 2002, Schreiber 2017). Clusters of cases were then mapped against a recent consensus statement of joint-related histopathological classification (Krenn et al. 2014). Between-group differences in clinical variables after HCA were compared using either Kruskal–Wallis test or chi-square test.

In HCA, our main interest was to find clusters of cases (hips) based on their dissimilarity. Since our data comprised binary and ordinal variables, we chose the Gower method to form the distance matrix (Gower 1971). After the selection of the appropriate dissimilarity measurement, clustering began by assigning each case to be an individual cluster forming a proximity matrix sized 284 columns × 284 rows. The matrix reflected the closeness of each cluster. Each case began as an individual cluster and was gradually merged with the most closely related cluster (of cases). We used the complete linkage method. This process was repeated until one single cluster remained. Our aim was to identify any meaningful and histologically relevant clusters. Hence, we did not use solely the agglomerative approach, which is the most commonly used method, to establish the optimal number of clusters. The agglomerative process uses the agglomeration schedule in which the change in agglomeration coefficient is depicted as the distance between merged clusters. The higher the change in the agglomeration coefficient, the higher is the dissimilarity between clusters. We interpreted the last stages of the clustering process to define the meaningful clusters of observations as five or less clusters were expected to be seen. "Natural break" was defined as the largest change in agglomeration coefficient producing meaningfully distinguishable clusters.

In addition to HCA, LCA was also performed to further analyze the possible underlying structures in our data set. LCA also aims to identify meaningful groups or class memberships of cases according to their (dis)similarity. To identify the optimal set of groups, LCA was first performed with two groups, then three groups, and so on. Akaikes Information Criterion (AIC) indices were interpreted to assess the most suitable baseline model. By using cluster analysis with LCA, we

aimed to have both the optimal model suggested by the indices and a meaningful set of groups so that each group could be readily labeled.

We further aimed to validate our primary outcome after cluster analysis and LCA by first running a validation analysis using both techniques and then separately for both implant groups. The rationale for this was the different wear behaviors between stemmed THAs and hip resurfacings. Bearing wear is seen in both implants, but taper corrosion is only seen in THA. If our segmentation techniques are robust against one major etiological factor, similar clusters and class memberships should be produced regardless of the implant type included in the analysis. Each cluster formed by validation clustering was matched against primary clusters. The distribution of cases among clusters obtained from validation cluster analysis was cross-tabulated against primary clustering to see whether discordant cases, that is, negative matches among two different clustering processes, existed. Validation LCA was performed in an equal way using the same principle as with the study cohort (all cases included).

In study IV, the statistical significance of the difference in wear volume between the higher and lower wearing side was tested using Wilcoxon signed ranks test (related samples). Mann-Whitney U-test was used to test the difference in wear volume distribution between the hips in patients with symmetric versus asymmetric histological and imaging findings (independent samples). The differences in histological findings between left and right hips were compared and the number of patients with identical findings, patients with a difference of one point and a difference of two points between the sides were calculated. The statistical significance of the difference in histological findings between the sides was tested with marginal homogeneity test except for the difference in presence of germinal centers, which did not fill the test requirements, and the McNemar test was used instead (Bonnini et al. 2014). Whether the presence of MRI-confirmed pseudotumor was similar between left and right sides was tested using McNemar test (related samples).

## 4.2.10   Ethical considerations

All studies were approved by the ethical committee of Pirkanmaa Hospital District (R11196 and R11006). All patients gave written informed consent to participate in the studies. Principles of Helsinki Declaration were strictly obeyed. No harm was done, or strain placed on patients. Participating in the studies did not affect the level or type of care patients received. Opting out from the study was possible at any time without negative consequences to the patient.

# 5 RESULTS

## 5.1 Wear of the ASR hip resurfacing (Study I)

In study I, 85 hips in 78 patients were analyzed for bearing wear volume. Analysis of the explanted components demonstrated a wide range of wear in both the acetabular cup and the femoral head (Table 9). Median total wear was 39 mm$^3$. Wear rates were also highly variable with a median of 9.0 mm$^3$/year (range, 1.1 - 99.7 mm$^3$/year). In a vast majority of the components (85.1%), the femoral head was more worn than the acetabular cup. Median ratio for head wear to cup wear was 1.7 (range, 0.5 - 10).

**Table 9.** Median volumetric wear and range for acetabular and femoral components and both combined.

| Component | Median volumetric wear (mm$^3$) | Range (mm$^3$) |
|---|---|---|
| Acetabular cup | 14 | 2 – 247 |
| Femoral head | 24 | 4 – 485 |
| Both combined | 39 | 7 – 541 |

## 5.2 Whole blood and synovial fluid metal ion levels and their association with bearing wear (Studies I and II)

In study I, in addition to actual volumetric component wear, WB and SF metal ion levels were also highly variable (Table 10). Cobalt was present in higher concentrations than chromium. In study II, we noted similarly variable concentrations. Subgroup analysis of the THAs and hip resurfacings separately revealed significantly higher concentrations of cobalt ions in THA compared with hip resurfacings (Table 11). We could not detect a difference in chromium concentrations between the groups.

**Table 10.** Median concentrations and ranges for chromium and cobalt ions in both whole blood and synovial fluid.

| Metal ion | Whole blood (µg/l) | Range (µg/l) | Synovial fluid (µg/l) | Range (µg/l) |
|-----------|--------------------|--------------| ----------------------|--------------|
| Chromium | 9.7 | 0.5 – 93.9 | 701 | 7.0 – 52360 |
| Cobalt | 15.4 | 0.7 – 224.7 | 281.5 | 27.0 – 14870 |

**Table 11.** Median values, p-values for comparison between groups and ranges for whole blood metal ion concentrations in patients with unilateral total hip arthroplasty (n=69) or hip resurfacing (n=13) patients.

| | Total hip arthroplasty | | | Hip resurfacing | | |
|-------|------------------------------------------------|--------------|-----------------------------------------------------|--------------|---------|
| Metal | Median concentration in whole blood (µg/l) | Range (µg/l) | Median concentration in whole blood (µg/l) | Range (µg/l) | P-value |
| Chromium | 3.7 | 0.4 - 29.9 | 3.9 | 1.5 - 7.2 | 0.60 |
| Cobalt | 11.0 | 0.6 - 108.5 | 3.9 | 1.5 - 16.2 | 0.001 |

In study I, the total wear volume of the head and cup strongly correlated with WB metal ion concentrations (Cr: $\varrho = 0.80$, $p < 0.001$ and Co: $\varrho = 0.84$, $p < 0.001$) and moderately with SF metal ion concentrations (Cr: $\varrho = 0.50$, $p < 0.01$ and Co: $\varrho = 0.41$, $p = 0.027$). Wear rate had slightly stronger correlation with WB metal ion concentrations (Cr: $\varrho = 0.87$, $p < 0.001$ and Co: $0.89$, $p < 0.001$) and SF metal ion concentrations (Cr: $\varrho = 0.71$, $p < 0.001$ and Co: $0.66$, $p < 0.01$) than total wear volume.

## 5.3 Metal debris accumulation in periprosthetic tissues and its relation to WB and SF metal ion levels (Study II)

Chromium had the highest concentration of all metals in the periprosthetic tissue in both the hip resurfacing and THA groups (Table 12). Titanium was elevated above the detection limit in nine patients with hip resurfacing and in 29 patients with THA. The concentrations for aluminum and vanadium did not reach the detection limit in any of the patients and were thus omitted from the analyses. We could not detect differences in periprosthetic tissue metal concentrations between THA and hip resurfacing groups (Table 12). In the tissue samples of the two

control patients, only the concentration of chromium exceeded the detection limit (0.3 µg/g and 0.5 µg/g, respectively).

**Table 12.** Median values with respected p-values and ranges for periprosthetic tissue metal concentrations in patients with total hip replacements (n = 87) and hip resurfacings (n = 20).

| Metal | Total hip arthroplasty | | Hip resurfacing | | |
|---|---|---|---|---|---|
| | Median concentration (µg/g) | Range (µg/g) | Median concentration (µg/g) | Range (µg/g) | P-value |
| Chromium | 39.2 | 0.4 - 1955.0 | 43.8 | 0.6 - 922.1 | 0.60 |
| Cobalt | 6.4 | 0.2 - 262.0 | 3.2 | 0.2 - 248.8 | 0.19 |
| Molybdenum | 1.8 | 0.2 - 174.6 | 0.5 | 0.2 - 32.4 | 0.080 |
| Titanium | 5.8 | 3.0 - 118.9 | 4.9 | 4.9 - 25.3 | 0.10 |

Periprosthetic tissue chromium and cobalt concentrations correlated weakly with whole blood and synovial fluid chromium and cobalt concentrations in the THA group (Table 13). In the resurfacing group, only periprosthetic tissue cobalt concentration reached statistically significant correlation with synovial fluid cobalt concentration (Table 13).

**Table 13.** Spearman rho correlation coefficients between tissue metal concentrations, whole blood (WB) and synovial fluid (SF) metal ion concentrations in total hip arthroplasty (n = 87) and hip resurfacing (n = 20) groups.

| | Total hip arthroplasty | | Hip resurfacing | |
|---|---|---|---|---|
| | Tissue chromium | Tissue cobalt | Tissue chromium | Tissue cobalt |
| WB chromium | rho= 0.32, p<0.01 | | rho= 0.48, p=0.10 | |
| WB cobalt | | rho= 0.31, p<0.01 | | rho= 0.24, p=0.43 |
| SF chromium | rho= 0.29, p<0.01 | | rho= 0.63, p=0.067 | |
| SF cobalt | | rho= 0.34, p<0.01 | | rho= 0.70, p=0.035 |

## 5.4     Histopathological findings in periprosthetic tissues (Studies I, II, III and IV)

In all studies, we noted wide variability between patients in histopathological presentation of the tissues. Pronounced inflammatory response (macrophages, lymphocytes, granulomas, germinal centers) and tissue necrosis were observed in vastly variable degrees. All tissues evinced at least mild macrophage infiltration (macrophage sheet thickness score of $\geq 1$), and in some cases strong infiltration was observed. Granulomas were observed in a minority of the samples. In regard to perivascular lymphocyte infiltration, most tissues evinced little to no lymphocytic cuffing (scores 0 and 1) and the cuffing was thick in only a few samples (scores 2 and 3). All cases with heavy lymphocyte cuffs had a macrophage sheet thickness score of 1, that is, there was only a little macrophage infiltration in these tissues (Study I). We noted the presence of germinal centers in only a small percentage of samples. The degree of necrosis was approximately evenly distributed in its four categories. In study I, the grade of necrosis correlated with the thickness of the perivascular lymphocyte cuff (rho = 0.41, p<0.001). Further, all cases with heavy lymphocytic infiltration displayed either grade 3 or 4 necrosis. All five tissue samples with germinal centers had grade 4 necrosis.

In studies II and III, we compared histopathological findings in THA versus hip resurfacing patients (Tables 14 and 15). In both studies, lymphocyte cuff thickness score was higher in patients with THA versus hip resurfacing and the difference was statistically significant. However, no difference was detected in macrophage sheet thickness between the hip resurfacing and THA groups. The grade of tissue necrosis was higher in the THA group compared with the hip resurfacing group. Furthermore, in Study III, we noted that extracellular metal was present more frequently in THAs versus hip resurfacings (Table 15).

**Table 14.**  Lymphocyte cuff thickness, macrophage sheet thickness and grade of necrosis in total hip arthroplasty group (n = 87) and hip resurfacing group (n = 20).

| | | Total hip arthroplasty | Hip resurfacing | P-value for group comparison |
|---|---|---|---|---|
| Lymphocyte cuff thickness | 0 (absent) | 33 (37.9%) | 15 (75%) | |
| | 1 (0-0.25 mm) | 41 (47.1%) | 4 (20%) | |
| | 2 (0.25-0.75 mm) | 13 (14.9%) | 1 (5.0%) | |
| | 3 (>0.75 mm) | 0 (0%) | 0 (0%) | |
| | | | | 0.011 |
| Macrophage sheet thickness | 0 (absent) | 1 (1%) | 0 (0%) | |
| | 1 (<1mm) | 68 (78.2%) | 18 (90%) | |
| | 2 (1-2mm) | 16 (18.4%) | 2 (10%) | |
| | 3 (>2mm) | 2 (2.3%) | 0 (0%) | |
| | | | | 0.65 |
| Grade of necrosis | 1 | 3 (3.4%) | 8 (40%) | |
| | 2 | 19 (21.8%) | 4 (20%) | |
| | 3 | 12 (13.8%) | 1 (5%) | |
| | 4 | 53 (60.9%) | 7 (35%) | |
| | | | | <0.001 |

**Table 15.**    Distribution of histological findings in 284 failed ASR metal-on-metal hip replacements. THA = total hip arthroplasty.

| Observation | | THA | | Hip resurfacing | | p-value for group comparison |
|---|---|---|---|---|---|---|
| | | Number | Percent | Number | Percent | |
| Plasma cells | Present | 48 | 23.1% | 13 | 16.9% | 0.3 |
| Diffuse lymphocytic inflammation | Present | 55 | 26.5% | 15 | 19.5% | 0.3 |
| Germinal center | Present | 12 | 5.8% | 6 | 7.8% | 0.6 |
| Lymphocyte cuff | Absent | 72 | 34.8% | 39 | 50.6% | 0.005 |
| | <0.25 mm | 112 | 54.1% | 32 | 41.6% | |
| | 0.25 – 0.75 mm | 23 | 11.1% | 4 | 5.2% | |
| | >0.75 mm | 0 | 0% | 2 | 2.6% | |
| Macrophage sheet | Absent | 6 | 2.9% | 0 | 0% | 0.4 |
| | <1 mm | 151 | 72.9% | 58 | 75.3% | |
| | 1-2 mm | 44 | 21.2% | 15 | 19.5% | |
| | >2 mm | 6 | 2.9% | 4 | 5.2% | |
| Granulomas | Present | 30 | 14.5% | 17 | 22.1% | 0.15 |
| Necrosis | Grade I | 7 | 3.3% | 21 | 27.3% | <0.001 |
| | Grade II | 50 | 24.2% | 22 | 28.6% | |
| | Grade III | 53 | 25.6% | 16 | 20.8% | |
| | Grade IV | 97 | 46.9% | 18 | 23.4% | |
| Extracellular metal | Present | 95 | 45.9% | 12 | 15.6% | <0.001 |
| Particle load | Absent | 44 | 21.2% | 13 | 16.9% | 0.6 |
| | Grade 1 | 40 | 19.3% | 13 | 16.9% | |
| | Grade 2 | 50 | 24.2% | 19 | 24.7% | |
| | Grade 3 | 49 | 23.7% | 18 | 23.4% | |
| | Grade 4 | 24 | 11.6% | 14 | 12.2% | |

## 5.5 Within-patient variability of histological, imaging and wear findings in patients with bilateral MoM hip replacements (Study IV)

As noted in the previous chapter, the between-patient variability of histological findings was high. In study IV, we investigated within-patient variability between contralateral hips in those MoM patients that had been revised bilaterally. Interestingly, we observed that the within-patient variability in histological findings was low. The congruence of histological findings between the left and the right hips is presented in Table 16. In the majority of the histological features and also in the majority of the patients, there were no differences between the hips ($p > 0.05$ for all comparisons). These features included macrophage sheet thickness, perivascular lymphocyte cuff thickness, presence of plasma cells, presence of diffuse lymphocytic infiltration and presence of germinal centers. In lymphocyte cuff thickness, the difference between the sides was at most 1 point. In macrophage sheet thickness, the findings were similar in 18 patients (62%), differed by 1 point in 9 patients (31%) and differed by 2 points in 2 patients (7%), respectively. The only histological findings that differed between the hips were grade of necrosis ($p < 0.01$) and presence of granulomas ($p = 0.025$). In the grade of necrosis there was a wide distribution in the difference between the sides. All patients had at least two histological variables with similar findings in both hips. The majority of patients (75.9%) had four or more histological variables with similar findings on both sides (Table 17). There were no differences in the similarity or dissimilarity of histological findings between left and right hips in males versus females (Table 18).

**Table 16.** Congruence in histological grading between contralateral (left and right) hips (within-subject).

| | Difference in histological grading between contralateral sides | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | No difference | 1 p | 2p | 3p | 4p | Scale |
| Macrophage sheet thickness | 18 (62%) | 9 (31%) | 2 (7%) | - | | 0-3 p |
| Lymphocyte cuff thickness | 14 (48%) | 15 (52%) | - | - | | 0-3 p |
| Degree of necrosis** | 6 (21%) | 10 (34%) | 9 (31%) | 1 (3%) | 3 (10%) | 0-4 p |
| Presence of plasma cells | 26 (90%) | 3 (10%) | | | | Yes/no |
| Presence of diffuse lymphocytic infiltration | 19 (66%) | 10 (34%) | | | | Yes/no |
| Presence of germinal centers | 27 (93%) | 2 (7%) | | | | Yes/no |
| Presence of granulomas** | 24 (83%) | 5 (17%) | | | | Yes/no |

Percentages represent proportion of all patients. In variables marked with **, there was a statistically significant ($p<0.05$) difference between the sides.

**Table 17.** The degree of similarity between the hips measured by the number of histological variables with similar findings on both sides in each patient.

| Histologic variables with symmetric findings on both sides | Number of patients | Percentage of patients |
|---|---|---|
| 1 | 0 | 0% |
| 2 | 2 | 6.9% |
| 3 | 5 | 17.2% |
| 4 | 6 | 20.7% |
| 5 | 6 | 20.7% |
| 6 | 9 | 31.0% |
| 7 | 1 | 3.4% |
| | Total 29 | Total 100 % |

**Table 18.** Comparison of similar and unsimilar histological findings between the sides in males versus females.

| Histological variable | Symmetric findings on both hips | Males | Females | P-value |
|---|---|---|---|---|
| Macrophage sheet thickness | Yes | 11 (69%) | 7 (54%) | |
| | No | 5 (31%) | 6 (46%) | 0.46 |
| Lymphocytic cuff thickness | Yes | 9 (56%) | 5 (38%) | |
| | No | 7 (44%) | 8 (62%) | 0.46 |
| Degree of necrosis | Yes | 4 (25%) | 2 (15%) | |
| | No | 12 (75%) | 11 (85%) | 0.66 |
| Presence of plasma cells | Yes | 14 (88%) | 12 (92%) | |
| | No | 2 (12%) | 1 (8%) | 0.58 |
| Presence of diffuse lymph. | Yes | 12 (75%) | 7 (54%) | |
| | No | 4 (25%) | 6 (46%) | 0.27 |
| Presence of germinal centers | Yes | 15 (94%) | 12 (92%) | |
| | No | 1 (6%) | 1 (8%) | 1.0 |
| Presence of granulomas | Yes | 15 (94%) | 9 (69%) | |
| | No | 1 (6%) | 4 (31%) | 0.14 |

Bilateral MRI classification for the presence of pseudotumors was available for 25 patients (86% of all patients). A total of 18 patients (72% of the classified) had either bilateral pseudotumors or no pseudotumors at all on either side, that is, the hips were symmetrical in regard to pseudotumor formation. There was no statistically significant difference in the presence of pseudotumor between the sides (p = 0.13). Of these 18 patients, 7 had pseudotumor on both sides (of which two were identical by exact classification) and 11 had no pseudotumor on either side.

Contrary to histological and imaging findings, there was a statistically and clinically significant difference between the contralateral sides in wear volume. Component wear was available bilaterally for 17 (59% of all) patients. Total wear volume in either hip ranged from 3 mm$^3$ to 94 mm$^3$ (median 13 mm$^3$, IQR 10 - 32 mm$^3$). The median difference in wear volume between higher and lower wearing side was 15.35 mm$^3$ (range 1 to 39 mm$^3$, IQR 6 - 23 mm$^3$) (p<0,001). This difference is illustrated in Figure 10. The median ratio of total wear volume between the hips was 2.0 (range 1.09 to 10.0, IQR 1.67 - 3.72). In 9 of the 17 (53%) patients with wear data available, the ratio of wear was 2.0 or greater, that is, there was at least a two-fold difference in the wear volume between the contralateral hips. Patients with asymmetrical pseudotumor finding between the sides evinced a similar distribution of total wear volume between the sides as those patients with symmetrical pseudotumor finding (Table 19). In addition, there were no differences in the total wear volumes of the hips in patients with pseudotumor on both sides (median 20.0 mm$^3$, range 9.0 to 111.0) versus no pseudotumor on either side (median 16.3 mm$^3$, range 3.0 to 51.0) (p = 0.28 for comparison).

**Table 19.**   The difference in total wear volume between contralateral sides in patients with symmetrical versus asymmetrical pseudotumor finding between the sides. Only patients with complete wear data are included (n=17).

|  | Pseudotumor finding between contralateral sides | | |
|---|---|---|---|
|  | Symmetrical | Asymmetrical | P-value |
| Median difference in total wear volume between the sides (mm$^3$) | 12.7 | 13.5 | 0.79 |
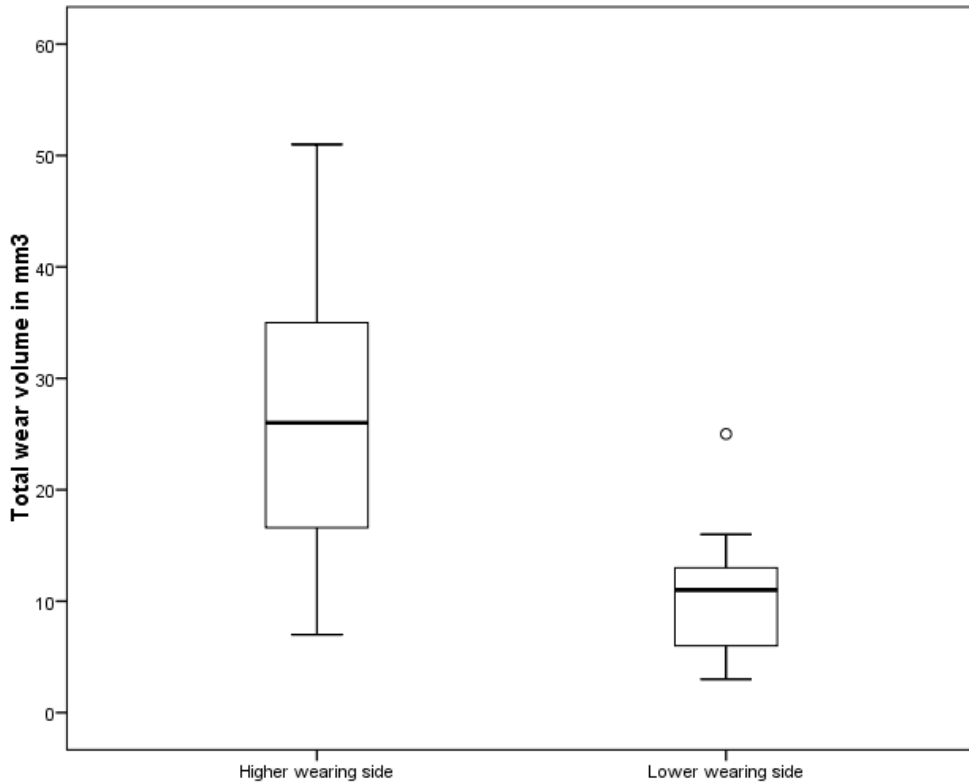
**Figure 10.** The difference in total wear volume between higher and lower wearing sides.

## 5.6 Comparison of two histopathological scoring methods (Study I)

In study I, tissues were scored using two different semiquantitative methods previously described in the literature – ALVAL and Natu grading (Campbell et al. 2010, Natu et al. 2012). We observed that perivascular lymphocyte cuff thickness (Natu grading) strongly correlated with inflammatory infiltrate score (ALVAL grading) ($\varrho = 0.79$, $p < 0.001$). Further, grade of necrosis (Natu grading) had strong correlation with synovial lining ($\varrho = 0.86$, $p < 0.001$) and tissue organization score ($\varrho = 0.80$, $p < 0.001$), which are components of the ALVAL score.

## 5.7    Latent histopathological subgroups observed (Study III)

In study III, HCA and LCA were performed to establish the underlying structure and relationships of the histological observations and to find similar cohorts of cases. Clustering and latent class analyses suggested four distinct histopathological subtypes that could be readily and reasonably labeled and mapped against a recent consensus statement (Krenn et al. 2014) (Tables 20 and 21). Cluster 1 in HCA and Class 1 in LCA could be readily labeled as "abrasion-induced foreign body Type I neosynovitis" (Figure 11). The characteristics of this subgroup were absence of necrosis or mild necrosis, and lack of diffuse synovitis, germinal centers, plasma cells and granulomas. Perivascular lymphocyte cuffs were also mainly absent. Macrophage sheets were mainly thin and particle load inside them was moderate. Cluster 2 in HCA and Class 2 in LCA were labeled as "abrasion-induced necrotic Type I neosynovitis" (Figure 12). As in the foreign body reaction, diffuse synovitis, plasma cells, germinal centers and perivascular lymphocytic cuffs were absent. However, granulomas were seen; macrophage sheets were thicker; the level of necrosis was moderate; the particle load inside macrophages was higher, and extracellular metals particles were present. Cluster 3 in HCA and Class 3 in LCA were similar. Plasma cells and germinal centers were prevalent; the level of necrosis was very high; lymphocytic cuffs were thick, and both particle load and extracellular metal content was low or absent. We labeled Cluster 3 in HCA and Class 3 in LCA as "immunologic Type IV neosynovitis" (Figure 13). The remaining subgroups, that is, Cluster 4 in HCA and Class 4 in LCA, were partly similar to the previous reaction, but plasma cells and germinal centers were less frequent and lymphocytic cuffs were thinner. For overview of common features with Class/Cluster 3, see Figure 13. However, in Class/Cluster 4 necrosis was moderate or high and also diffuse synovitis was most frequent. Extracellular metal particles were commonly present and particle load within macrophages was high. Hence, these subgroups were labeled as "abrasion induced inflammatory lymphocytic Type I neosynovitis." A summary of the characteristics of each subgroup is shown in Table 22.
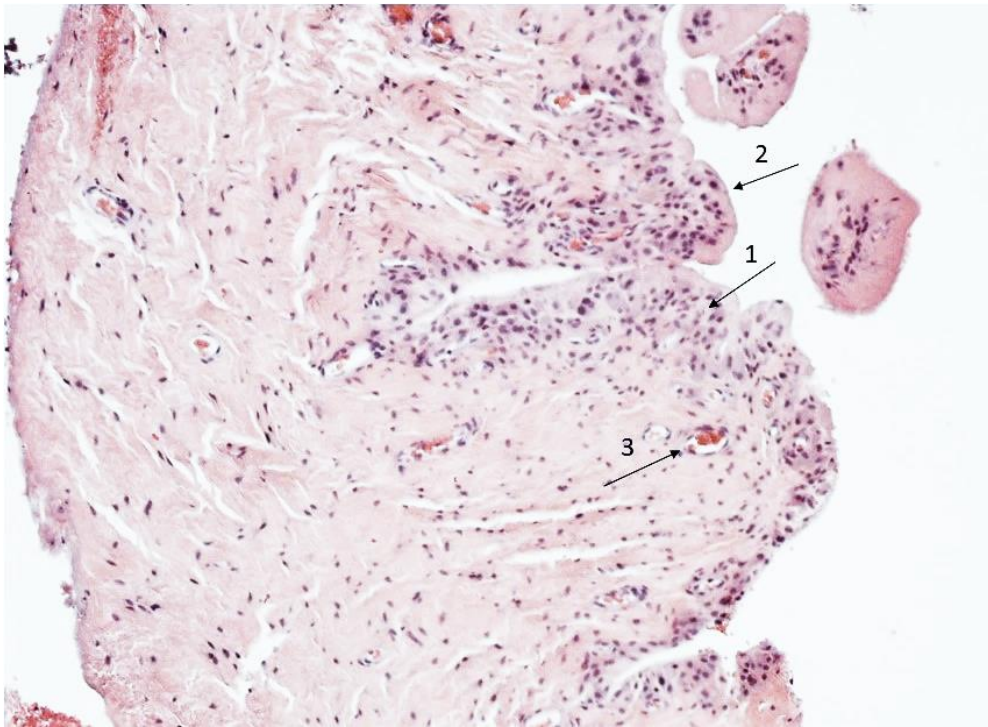
**Figure 11.** Photomicrograph of synovial tissue classified as Cluster/Class 1. Arrow 1 points to subintimal macrophage sheet. Arrow 2 shows synovial lining which is mostly intact. Arrow 3 shows a blood vessel. No visible necrosis or lymphocyte accumulations. Captured with Nikon Eclipse 50i light microscope fitted with 20x objective (total magnification 200x).

**Figure 12.** Photomicrograph of synovial tissue classified as Cluster/class 2. Arrow 1 points to disrupted synovial lining. Arrow 2 shows a thick subintimal macrophage sheet. Arrow 3 points to macrophages with intracellular metallic debris. Arrow 4 points to necrotic acellular tissue. Captured with Nikon Eclipse 50i light microscope fitted with 20x objective (total magnification 200x).
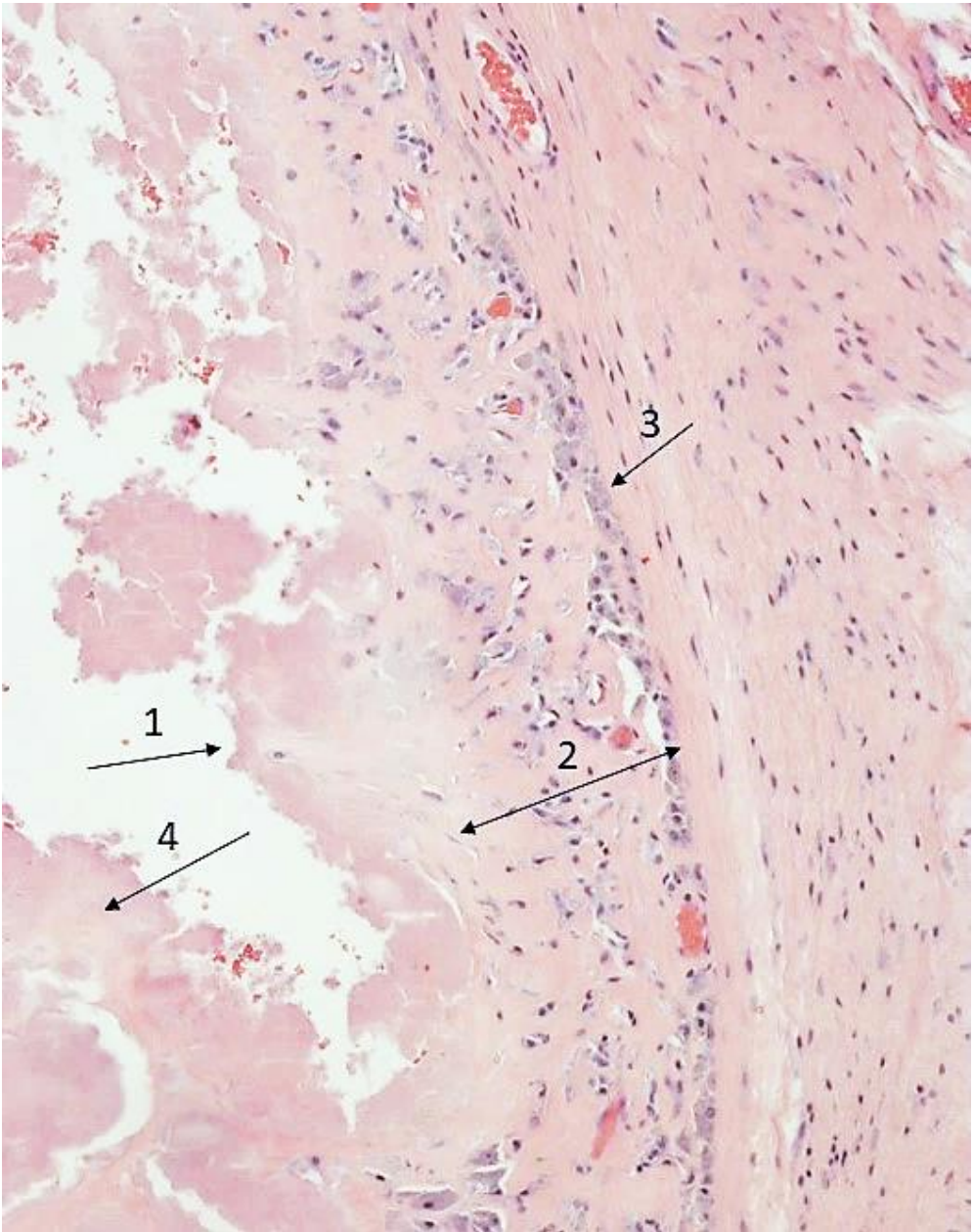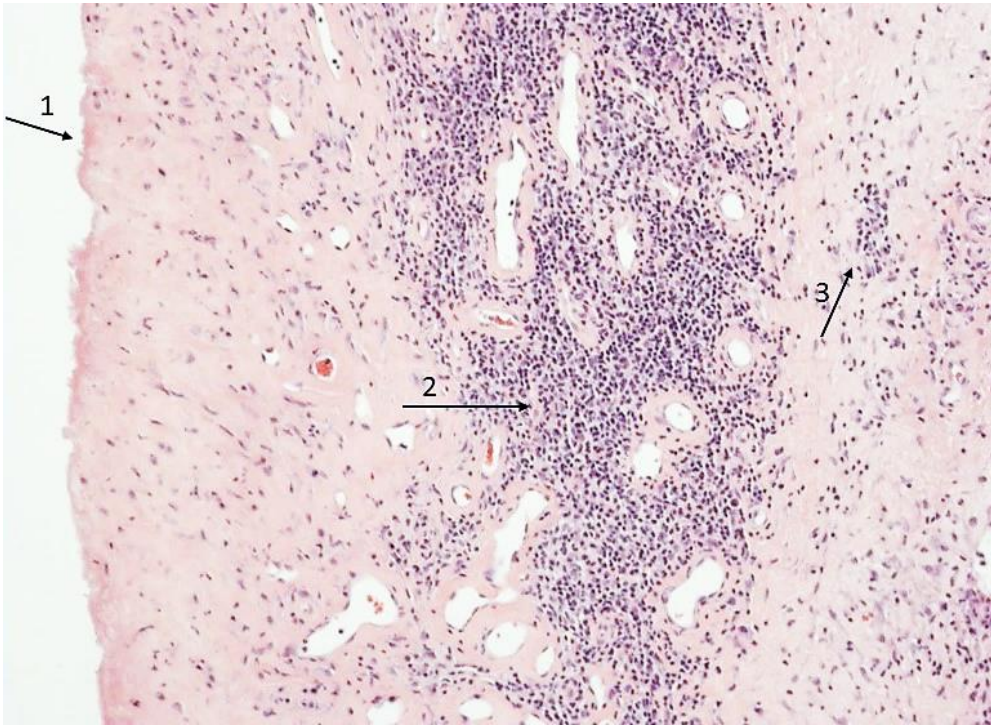
**Figure 13.** Photomicrograph of synovial tissue classified as Cluster/class 3. Arrow 1 represents disrupted synovial lining. Arrow 2 shows thick perivascular lymphocyte cuffs which are interconnected. Arrow 3 demonstrates diffuse lymphocytes. Captured with Nikon Eclipse 50i light microscope fitted with 20x objective (total magnification 200x).

**Table 20.** Distribution of histological findings divided by the four groups or clusters resulting from cluster analysis.

| Observation | | Cluster 1 Abrasion-induced foreign body type I neosynovitis | Cluster 2 Abrasion-induced necrotic Type I neosynovitis | Cluster 3 Immunologic type IV neosynovitis | Cluster 4 Abrasion-induced inflammatory lymphocytic type I neosynovitis |
|---|---|---|---|---|---|
| Number of hips | | 68 hips | 78 hips | 88 hips | 50 hips |
| Diffuse lymphocytic inflammation present | | 0.0% | 0.0% | 25% | 95.0% |
| Germinal center present | | 0.0% | 0.0% | 15.9% | 8.0% |
| Plasma cells present | | 1.3% | 12.0% | 55.7% | 12.0% |
| Granulomas present | | 0.0% | 28.2% | 20.5% | 14.0% |
| Lymphocytic cuff | Absent | 72.1% | 55.1% | 17.0% | 8.0% |
| | <0.25 mm | 27.9% | 42.3% | 61.4% | 76.0% |
| | 0.25 - 0.75 mm | 0.0% | 2.6% | 19.3% | 16.0% |
| | >0.75 mm | 0.0% | 0.0% | 2.3% | 0.0% |
| Macrophage sheets | Absent | 1.5% | 1.3% | 3.4% | 2.0% |
| | <1 mm | 75.0% | 64.1% | 76.1% | 82.0% |
| | 1-2 mm | 20.6% | 26.9% | 18.2% | 16.0% |
| | >2 mm | 2.9% | 7.7% | 2.3% | 0.0% |
| Necrosis | Grade I | 29.4% | 7.7 % | 0.0% | 4.0% |
| | Grade II | 63.3% | 25.6 % | 2.3% | 14.0% |
| | Grade III | 7.4% | 35.9 % | 20.5% | 36.0% |
| | Grade IV | 0.0% | 30.8 % | 77.3% | 46.0% |
| Extracellular metal content present | | 0.0% | 82.1% | 26.1% | 40.0% |
| Particle load | Absent | 17.6% | 5.1% | 31.8% | 26.0% |
| | Grade 1 | 14.7% | 21.8% | 23.9% | 10.0% |
| | Grade 2 | 32.4% | 20.5% | 26.1% | 16.0% |
| | Grade 3 | 22.1% | 32.1% | 11.4% | 34.0% |
| | Grade 4 | 13.2% | 20.5% | 6.8% | 14.0% |

**Table 21.**  Distribution of histological findings divided by the four groups or classes resulting from latent class analysis.

| Observation | | Class 1 Abrasion-induced foreign body type I neosynovitis | Class 2 Abrasion-induced necrotic Type I neosynovitis | Class 3 Immunologic type IV neosynovitis | Class 4 Abrasion-induced inflammatory lymphocytic type I neosynovitis |
|---|---|---|---|---|---|
| Number of hips | | 116 hips | 41 hips | 37 hips | 90 hips |
| Germinal center present | | 0.0% | 0.0% | 44.1% | 1.6% |
| Diffuse lymphocytic inflammation present | | 4.2% | 9.7% | 40.0% | 51.8% |
| Plasma cells present | | 0.0% | 7.7% | 47.2% | 44.7% |
| Granulomas present | | 5.5% | 31.3% | 27.2% | 19.0% |
| Lymphocytic cuff | Absent | 69.9% | 70.0% | 4.1% | 0.0% |
| | <0.25 mm | 31.6% | 27.3% | 52.7% | 87.1% |
| | 0.25-0.75 mm | 0.% | 2.7% | 40.6% | 11.8% |
| | >0.75 mm | 0.0 % | 0.0% | 2.6% | 1.1% |
| Macrophage sheets | Absent | 0.0% | 3.1% | 12.4% | 0.0% |
| | <1 mm | 87.5% | 6.9% | 75.7% | 86.9% |
| | 1-2 mm | 11.4% | 74.3% | 6.6% | 13.1% |
| | >2 mm | 1.1% | 15.6% | 5.3% | 0.0% |
| Necrosis | Grade I | 24.5% | 0.0% | 0.0% | 0.0% |
| | Grade II | 43.6% | 28.5% | 9.8% | 7.7% |
| | Grade III | 15.1% | 33.3% | 0.0% | 41.1% |
| | Grade IV | 17.3% | 38.3% | 90.2% | 51.2% |
| Extracellular metal content present | | 29.3% | 40.5% | 0.0% | 62.9% |
| Particle load | Absent | 6.7% | 6.2% | 59.7% | 27.1% |
| | Grade 1 | 20.7% | 14.9% | 11.7% | 21.0% |
| | Grade 2 | 33.7% | 16.6% | 22.1% | 16.6% |
| | Grade 3 | 28.5% | 18.6% | 3.1% | 29.5% |
| | Grade 4 | 10.3% | 43.7% | 3.5% | 5.9% |

**Table 22.**   Semi-qualitative descriptions of the four groups of synovial histopathological responses seen in ARMD.

| Abrasion-induced foreign body type I neosynovitis | Abrasion-induced necrotic Type I neosynovitis | Immunologic type IV neosynovitis | Abrasion-induced inflammatory lymphocytic type I neosynovitis |
|---|---|---|---|
| **Cluster/Class I** | **Cluster/Class II** | **Cluster/Class III** | **Cluster/Class IV** |
| Intact or mildly destructed synovia | Moderately or highly destructed synovia | Extremely destructed synovia | Highly destructed synovia |
| No or mild diffuse lymphocytic synovitis | No or mild diffuse lymphocytic synovitis | Diffuse lymphocytic synovitis present | Diffuse lymphocytic synovitis present |
| No plasma cells | No plasma cells | Plasma cells clearly present | Plasma cells may be present |
| No germinal centers | No germinal centers | Germinal centers present | No germinal centers |
| Perivascular lymphocyte cuffs mostly absent | Perivascular lymphocyte cuffs absent | Perivascular lymphocyte cuffs thick | Perivascular lymphocyte cuffs present |
| Large extracellular metal particles may be present | Large extracellular metal particles clearly present | Large extracellular metal particles absent | Large extracellular metal particles may be present |
| Moderate to high number of metal particles inside histiocytes | Moderate to high number of metal particles inside histiocytes | None or few metal particles inside histiocytes | Some metal particles inside histiocytes |
| Thin or moderately thick macrophage sheets | Moderately to very thick macrophage sheets | Macrophage sheets absent or thin | Thin macrophage sheets |

## 5.8 Comparison of MRI findings and metal ion levels (WB and SF) across different histopathological subgroups (Study III)

In study III, we found that the hips in both the "immunologic neosynovitis" and the "abrasion-induced lymphocytic neosynovitis" groups were prominently THAs, and that they had the lowest median levels of metals present in both SF and WB (Table 23). The lowest WB and SF cobalt levels were seen in the "immunologic neosynovitis" group. Thick-walled pseudotumors were most common in the "foreign body neosynovitis" and "inflammatory lymphocytic neosynovitis" groups. The highest metal levels in WB and SF were seen in the "abrasion-induced foreign body neosynovitis" and "abrasion-induced necrotic neosynovitis" groups. Hip resurfacings were most frequently present in the "abrasion-induced foreign body neosynovitis" group.

**Table 23.** Comparison of clinical variables across different subgroups. * p-value for rank comparison.

| | | Abrasion-induced foreign body type I neosynovitis | Abrasion-induced necrotic Type I neosynovitis | Immunologic type IV neosynovitis | Abrasion-induced inflammatory lymphocytic type I neosynovitis | P-value |
|---|---|---|---|---|---|---|
| Implant type | Resurfacing | 35 (51.5%) | 17 (21.8%) | 17 (19.3%) | 9 (18%) | <0.0001 |
| | Total hip | 33 (48.5%) | 61 (78.2%) | 71 (80.7%) | 41 (82%) | |
| Pseudotumor | No | 29 (49.2%) | 58 (70.7%) | 29 (58%) | 20 (45.5%) | 0.0012 |
| | Thin walled, cystic | 9 (15.3%) | 11 (13.4%) | 15 (30%) | 7 (15.9%) | |
| | Thick walled cystic, solid | 21 (35.6%) | 13 (15.9%) | 6 (12%) | 17 (38.6%) | |
| WB Cobalt (µg/l) | Median (IQR) | 13.6 (8.2-37.7) | 12.5 (7.9-19.4) | 8.3 (1.83-18.3) | 10.2 (5.4-15.6) | 0.0019* |
| Synovial Cobalt (µg/l) | Median (IQR) | 1595 (593-2772) | 1193 (530-1983) | 416 (226-1029) | 728 (556-1423) | 0.0056* |
| WB Chrome (µg/l) | Median (IQR) | 5.6 (2.4-12.8) | 4.2 (2.6-7.8) | 3.4 (1.93-10.1) | 3.1 (2.3-5.2) | 0.088* |
| Synovial Chrome (µg/l) | Median (IQR) | 872 (384-2870) | 992 (417-4293) | 447 (132-1883) | 652 (168-1473) | 0.11* |

## 5.9 Associations between wear, indirect measures of wear (WB, SF, tissue metals) and histopathological findings (Studies I and II)

### 5.9.1 Wear volume and volumetric wear rate (Study I)

In study I, associations between volumetric wear measurements and histopathological findings were investigated. Correlations between histological variables, total wear volume of the head and cup components combined, and wear rate are listed in Table 24. Total wear volume correlated with macrophage sheet thickness, grade of necrosis (Figure 14), synovial lining score, tissue organization score and total ALVAL score. Wear rate had similar correlations, but a correlation with macrophages could not be established (p = 0.069). Median wear rate for those tissues with heavy lymphocyte infiltration (score ≥2) and high degree of necrosis (grade ≥3) was lower than for those tissues with a lower number of lymphocytes observed (Table 25).
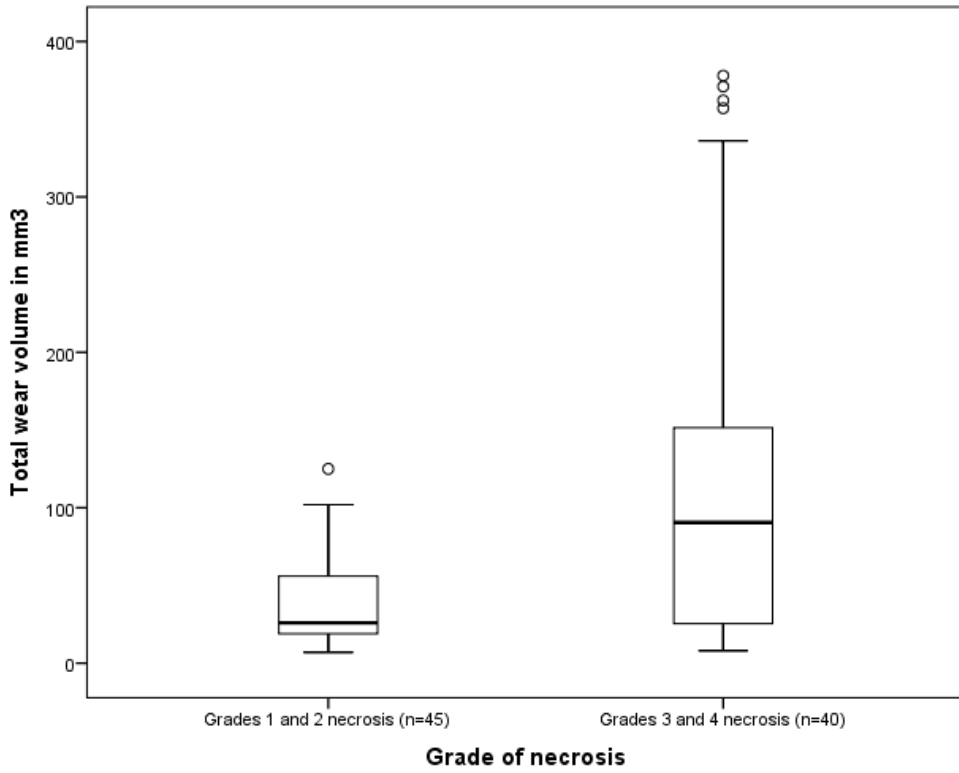
**Figure 14.** The difference in median total wear volume (head and cup) in patients with low-grade necrosis (grades 1 and 2) versus patients with high-grade necrosis (grades 3 and 4), p<0.001.

**Table 24.** Spearman rho correlation coefficients and associated p-values for correlations between total wear volume, wear rate, indirect markers of wear (whole blood and synovial fluid metal ion concentrations) and histopathological grading (Natu and ALVAL). Values that are statistically significant are flagged. WB = Whole Blood, SF = Synovial Fluid, Co = cobalt, Cr = chromium.

| | Natu grading | | | ALVAL grading | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Inflammatory infiltrate score | Synovial lining score | Tissue organization score | Total ALVAL score |
| Total wear volume | rho= 0.11 | rho= 0.25* | rho= 0.35* | rho= 0.13 | rho= 0.37* | rho= 0.25* | rho= 0.31* |
| | p= 0.32 | p= 0.020 | p< 0.01 | p= 0.23 | p< 0.01 | p= 0.023 | p< 0.01 |
| Wear rate | rho= 0.17 | rho= 0.20 | rho= 0.42* | rho= 0.16 | rho= 0.48* | rho= 0.35* | rho= 0.40* |
| | p= 0.12 | p= 0.069 | p< 0.0001 | p= 0.15 | p< 0.0001 | p<0.01 | p<0.001 |
| WB Cr | rho= 0.089 | rho= 0.30* | rho= 0.45* | rho= 0.19 | rho= 0.54* | rho= 0.33* | rho= 0.48* |
| | p= 0.51 | p= 0.024 | p< 0.001 | p= 0.26 | p< 0.001 | p= 0.015 | p< 0.001 |
| WB Co | rho= 0.18 | rho= 0.29* | rho= 0.51* | rho= 0.26 | rho= 0.60* | rho= 0.40* | rho= 0.55* |
| | p= 0.18 | p= 0.029 | p< 0.001 | p= 0.055 | p< 0.001 | p< 0.01 | p< 0.001 |
| SF Cr | rho= 0.30 | rho= 0.16 | rho= 0.48* | rho= 0.25 | rho= 0.56* | rho= 0.34 | rho= 0.53* |
| | p= 0.12 | p= 0.40 | p< 0.01 | p= 0.19 | p< 0.01 | p= 0.070 | p< 0.01 |
| SF Co | rho= 0.49* | rho= 0.17 | rho= 0.54* | rho= 0.44* | rho= 0.47* | rho= 0.38* | rho= 0.57* |
| | p< 0.01 | p= 0.37 | p< 0.01 | p= 0.017 | p= 0.011 | p= 0.045 | p< 0.01 |

**Table 25.**     Wear rates according to lymphocyte cuff thickness.

|  | Lymphocyte cuff < 2 | Lymphocyte cuff ≥ 2 | P-value |
|---|---|---|---|
| Median wear rate (mm³/year) | 8.1 | 14.5 | 0.054 |
| Range (mm³/year) | 1.1 - 99.8 | 3.7 - 48.0 |  |

## 5.9.2   Whole blood metal ion levels (Studies I and II)

In Study I, in a similar manner to wear volume and wear rate, WB cobalt and chromium ion levels correlated moderately with macrophage sheet thickness, grade of necrosis, synovial lining score, tissue organization score and total ALVAL score (Table 24). However, in Study II, only WB cobalt and metal particle load correlated with each other. No other correlations could be established between WB metal ion levels and any of the histological variables (Table 26) in either the THA or the hip resurfacing groups in Study II. Moreover, lymphocyte cuff thickness did not correlate with any of the wear or metal measurements in Studies I and II.

**Table 26.** Correlations between histological findings, periprosthetic tissue metal concentrations, whole blood metal ion levels (WB) and synovial fluid (SF) metal ion levels in the total hip replacement group (n = 87) and the hip resurfacing group (n = 20). Statistically significant correlations are flagged.

| | Total hip arthroplasty | | | | Hip resurfacing | | | |
|---|---|---|---|---|---|---|---|---|
| | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Metal particle load | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Metal particle load |
| Tissue chromium | rho= -0.20 p= 0.063 | rho= 0.022 p= 0.84 | rho= -0.13 p= 0.22 | rho= 0.34* p < 0.01 | rho= -0.36 p= 0.12 | rho= 0.12 p= 0.63 | rho= -0.28 p= 0.23 | rho= 0.29 p= 0.22 |
| Tissue cobalt | rho= -0.072 p= 0.51 | rho= 0.031 p= 0.78 | rho= -0.001 p= 0.99 | rho= 0.30* p < 0.01 | rho= -0.35 p= 0.13 | rho= 0.09 p= 0.72 | rho= -0.28 p= 0.23 | rho= 0.26 p= 0.27 |
| Tissue molybdenum | rho= -0.071 p= 0.514 | rho= 0.060 p= 0.584 | rho= -0.069 p= 0.53 | rho= 0.25* p= 0.019 | rho= -0.29 p= 0.21 | rho= 0.03 p= 0.90 | rho= -0.34 p= 0.15 | rho= 0.30 p= 0.20 |
| Tissue titanium | rho= -0.017 p= 0.88 | rho= -0.036 p= 0.74 | rho= -0.035 p= 0.74 | rho= 0.11 p= 0.30 | rho= -0.16 p= 0.51 | rho= -0.030 p= 0.60 | rho= -0.13 p= 0.58 | rho= -0.077 p= 0.75 |
| WB Cr | rho= -0.092 p= 0.45 | rho= 0.043 p= 0.73 | rho= 0.011 p= 0.92 | rho= 0.21 p =0.085 | rho= -0.34 p= 0.25 | rho= 0.29 p= 0.35 | rho= 0.14 p= 0.66 | rho= 0.32 p= 0.29 |
| WB Co | rho= -0.088 p= 0.47 | rho=-0.053 p= 0.67 | rho= 0.10 p= 0.41 | rho= 0.39* p < 0.01 | rho= -0.11 p= 0.72 | rho= 0.29 p= 0.35 | rho= 0.33 p= 0.27 | rho= 0.53 p= 0.067 |
| SF Cr | rho= -0.096 p= 0.39 | rho= 0.020 p= 0.86 | rho= -0.077 p= 0.49 | rho= 0.15 p= 0.18 | rho= -0.46 p= 0.22 | rho= 0.00 p= 1.00 | rho= 0.11 p= 0.78 | rho= 0.77* p= 0.016 |
| SF Co | rho= 0.053 p= 0.64 | rho= 0.12 p=0.30 | rho= 0.17 p= 0.12 | rho= 0.17 p= 0.14 | rho= -0.43 p= 0.25 | rho= 0.21 p= 0.59 | rho= 0.19 p= 0.62 | rho= 0.59 p= 0.096 |

## 5.9.3 Synovial fluid metal ion levels (Studies I and II)

In study I, SF chromium ion concentration correlated with grade of necrosis, synovial lining score and total ALVAL score (Table 24). SF cobalt ion concentration correlated with all histological variables – lymphocytic cuffing, macrophage sheet thickness, grade of necrosis, inflammatory infiltrate score, synovial lining score, tissue organization score and total ALVAL score (Table 24). Contrarily, in Study II, only SF chromium correlated with metal particle load in the hip resurfacing group. Neither SF chromium nor cobalt ion levels correlated with any other histological variables (Table 26).

## 5.9.4 Periprosthetic tissue metal concentrations (Study II)

In study II, neither periprosthetic chromium, cobalt, molybdenum nor titanium concentrations correlated with any of the histological variables except metal particle load (Table 26). However, in the THA group, median chromium concentration was lower in those tissues with lymphocytic cuffing present versus tissues with no lymphocytic cuffing at all (p= 0.045, Table 27). In regard to cobalt and molybdenum, no differences could be established (p>0.05). In the hip resurfacing group, concentrations of chromium and cobalt were lower in those tissues with lymphocytes present, but these differences were not statistically significant (Table 28).

**Table 27.**     Median metal concentration in tissues with lymphocytes present and tissues with no lymphocytes present in the total hip replacement group (n = 87).

| Median concentration in tissue (µg/g) | Lymphocytes present | No lymphocytes present | P-value |
|---|---|---|---|
| Chromium | 30.1 | 67.4 | 0.045 |
| Cobalt | 6.4 | 6.1 | 0.43 |
| Molybdenum | 1.7 | 1.8 | 0.38 |

**Table 28.** Median metal concentration in tissues with lymphocytes present and tissues with no lymphocytes present in the hip resurfacing group (n = 20).

| Median concentration in tissue (µg/g) | Lymphocytes present | No lymphocytes present | P-value |
|---|---|---|---|
| Chromium | 8.0 | 79.3 | 0.11 |
| Cobalt | 1.2 | 4.2 | 0.12 |
| Molybdenum | 0.3 | 0.69 | 0.20 |

# 6 DISCUSSION

## 6.1 Principal findings

This dissertation focused on the histopathological findings of soft tissues retrieved from patients with revised MoM hip replacements. We evaluated the role of latent histopathological subtypes, patient-related factors, clinical variables, implant wear, metal concentrations in tissues, SF and WB on the pathogenesis of the adverse reactions related to metal wear debris frequently seen in patients with MoM hip replacements.

Overall, the histopathological findings were vastly diverse. Inflammatory changes were present in all tissues. Notably, variable amounts of macrophages, lymphocytes, plasma cells, germinal centers, granulomas and necrosis were seen. We identified four different histopathological subtypes of ARMD – wear-related foreign-body macrophage response, wear-related necrotic response, immunologic lymphocytic adaptive response and wear-related lymphocytic response. Patients with THAs presented with the highest amounts of perivascular lymphocytes and tissue necrosis compared with patients with hip resurfacings. Implant wear and WB metal ion concentrations were associated with macrophage sheet thickness and grade of necrosis, but not lymphocyte cuff thickness. Tissue metal concentrations had poor or non-existent correlations with any of the histopathological variables except metal particle load in macrophages. In bilateral patients, we observed strong agreement in histopathological findings and imaging findings between contralateral sides despite markedly different wear volumes.

## 6.2 Direct wear measurements, indirect estimates of wear and their associations

The median total wear volume of the ASR hip resurfacing in our patient cohort was 39 mm$^3$. Park et al. reported a lower median total wear of 21 mm$^3$ for 21 ASR hip resurfacings (Park et al. 2018). Lord et al., on the other hand, reported a mean

volumetric wear of 29 mm$^3$ for 22 ASR acetabular cups and 22 mm$^3$ for 32 ASR femoral heads (Lord et al. 2011). Conversely, a simulator study of ASR resurfacings reported volumetric wear of approximately 5 mm$^3$ at 15 million cycles, roughly equivalent to 15 years of prosthesis use (Leslie et al. 2008). Wear of the ASR hip resurfacing is much higher than what would have been anticipated from the results of the simulator study. Hence, these results highlight the importance of retrieval studies to establish the true performance of hip replacements.

WB metal ion levels had strong correlation with bearing wear volume and wear rate. SF metal concentrations also correlated with bearing wear volume and wear rate, but the correlations were weaker. Wear rate had a stronger correlation with WB and SF metal ion levels than total bearing volume. Wear rate, WB and SF metal ion levels likely reflect the average recent burden of metal debris, whereas total wear volume reflects the amount of total wear accumulated during implantation time. Similarly to our study, De Smet et al. found that both WB and SF metal ion levels correlated well with linear wear of the femoral component (De Smet et al. 2008). Langton et al. also noted a strong correlation between wear volume and WB metal ion levels (Langton et al. 2011b). Measurement of WB metal ion levels is a reliable, indirect way to gain information of the in situ wear process. On the other hand, the measurement of SF metal concentrations does not seem to offer any additional information compared with WB measurement. Furthermore, aspiration of SF is more invasive. We therefore recommend using WB metal ion levels as a surrogate measure of implant wear.

## 6.3 Periprosthetic tissue metal concentrations and their associations with WB and SF metal concentrations

Chromium had the highest concentrations in periprosthetic tissues in both THAs and hip resurfacings. This is in line with previous research (Catelas et al. 2006, Hart et al. 2010, Lohmann et al. 2013, Scharf et al. 2014). The median concentrations of chromium in the periprosthetic tissue exceeded those of cobalt by more than six-fold in both study groups. Chromium is known to accumulate in the tissues to a high degree, whereas cobalt ions are rapidly transported to the blood and eliminated in the urine, which explains why chromium concentration is higher than cobalt in periprosthetic tissues (Merritt et al. 1989, Brown et al. 1993).

Periprosthetic metal concentrations correlated poorly with whole blood and synovial fluid metal ion concentrations. The only exception was the good

correlation between synovial fluid and periprosthetic tissue cobalt concentrations. Witt et al. also found no correlation between serum and periprosthetic metal ion levels (Witt et al. 2014). We suggest that the overall poor correlations are due to tissues reflecting the accumulated metal load, while whole blood and synovial fluid reflect the amount of wear that has been generated more recently. This is supported by Kuba et al. who noted that tissue concentrations are dependent on the in situ time of the implant (Kuba et al. 2019).

## 6.4 Histopathological findings and semiquantitative scoring methods

In all studies, we observed variable degrees of macrophage infiltration, perivascular and diffuse lymphocytes, necrosis, occasional plasma cells, granulomas and germinal cells. Similar descriptive findings have been made in a plethora of previous studies (Davies et al. 2005, Willert et al. 2005, Mahendra et al. 2009, Campbell et al. 2010, Natu et al. 2012, Grammatopoulos et al. 2013, Lohmann et al. 2013). We found a similar correlation between perivascular lymphocytic cuffing and necrosis as reported in a previous study (Langton et al. 2011b). Furthermore, all tissue samples with germinal centers presented with severe necrosis. Mittal et al. suggested that these tissues with germinal centers represent a pathological entity, whereas Natu et al. proposed that they are the end-stage of the continuum of lymphoid neogenesis (Natu et al. 2012, Mittal et al. 2013).

We compared two different scoring methods to assess periprosthetic tissues: the scoring system published by Natu et al. and ALVAL-scoring by Campbell et al. (Campbell et al. 2010, Natu et al. 2012). Both of these scoring methods have been used in several recent studies, but to the best of our knowledge no direct comparisons have been made (Campbell et al. 2018b). ALVAL grading is relatively restricted compared to Natu grading as it only includes inflammatory infiltrate subscore, synovial lining subscore and tissue organization subscore. In comparison, the Natu score is more comprehensive. It includes separate scores for thickness of macrophage sheets, lymphocyte cuffs, presence of diffuse lymphocytes, presence of germinal centers, granulomas and plasma cells, and grade of necrosis. However, as is the case with other scoring methods, the Natu method is still semiquantitative. Furthermore, the number of diffuse lymphocytes is not graded, only whether they are present or not.

As discussed by Ricciardi et al., both the synovial lining and tissue organization subscores of ALVAL grading reflect the degree of necrosis (Ricciardi et al. 2016). A strong correlation between these ALVAL subscores and the Natu score for necrosis was observed in our study. ALVAL score was originally designed to help distinguish failures related to high wear from failures related to low wear and suspected hypersensitivity response. Although necrosis is often seen with ALVAL response, it is not specific for ALVAL as it is also seen with macrophage-dominated reactions with possible related cytotoxicity (Mahendra et al. 2009, Grammatopoulos et al. 2013). This leaves only the inflammatory infiltrate score in ALVAL grading specific for ALVAL response. In the present study, a strong correlation between lymphocyte cuff thickness (Natu) and inflammatory infiltrate subscore (ALVAL) was observed. This indicates that the inflammatory infiltrate subscore is useful in distinguishing lymphocyte-dominated responses from those that are not lymphocyte-dominated. Phillips et al. also concluded that ALVAL scoring is useful for distinguishing between macrophage and lymphocyte responses (Phillips et al. 2014). Inflammatory infiltrate score involves the evaluation of both lymphocytic and macrophagic components. However, both lymphocytes and macrophages are often seen in periprosthetic tissues. Grammatopoulos et al. suggested that an easier method to identify ALVAL responses from wear-related responses would be to measure only the thickness of the lymphocytic cuffing, referred to as Oxford ALVAL score in their study (Grammatopoulos et al. 2013). We agree with Grammatopoulos et al. and find that separate scores for the evaluation of macrophage and lymphocyte infiltration provide more information about the failure mechanism and lead to easier comparisons between histopathological studies.

## 6.5    Latent subtypes of ARMD and related clinical findings

The results of the hierarchical cluster and LCA implied four distinct subtypes of histopathological findings in failed ASR MoM hip replacements. Our results suggest that the traditional and often acclaimed "ALVAL-type" responses may be present in two different histological entities, a finding that coincides with the recent consensus statement. These subtypes are naturally very similar in nature – both evince perivascular lymphocyte cuffs and a high grade of necrosis. Furthermore, both diffuse lymphocyte infiltration and plasma cells are commonly seen. The major difference between these subtypes, based on the results of the current study,

is the particle load within macrophages and extracellular metal particles. One subtype lacked extracellular metal particles, and particle load within histiocytes was low or absent. Further, metal concentrations in WB and SF were lowest in this group. This subtype also evinced a very high grade of synovial necrosis, and germinal centers were more common in this subtype than in other subtypes. We therefore suggest that this subtype represents the "immunologic Type IV neosynovitis" reaction, that is, a true hypersensitivity type ALVAL reaction. On the other hand, the other "abrasion-induced inflammatory lymphocytic Type I neosynovitis" can be regarded as an "ALVAL-type response" associated with wear. In this subtype, the histopathological findings are slightly milder compared with "immunologic type IV neosynovitis." The other groups in our analysis that coincide with the consensus statement were "abrasion-induced foreign body Type I neosynovitis" and "abrasion-induced necrotic Type I neosynovitis". These two histopathological entities have been proposed previously (Mahendra et al. 2009, Grammatopoulos et al. 2013).

It is noteworthy that hips in the "immunologic Type IV neosynovitis" group evinced the lowest levels of WB and SF cobalt and SF chromium despite the majority of hips being THAs in this group. This bodes well with the suspected hypersensitivity in this group of patients. In general, THAs produce higher metal ion levels compared with hip resurfacings (Lainiala et al. 2016). Moreover, THAs have an additional source of metal debris – taper wear and corrosion – this debris being different from the debris originating from the bearing surfaces (Xia et al. 2017). Based on their results, Xia et al. suggested that taper wear/corrosion debris is more immunogenic than bearing wear debris. This would explain why most hips in both the "immunologic type IV neosynovitis" and "abrasion-induced inflammatory lymphocytic neosynovitis" groups were THAs.

The current literature on the correlation between metal wear debris and ALVAL/lymphocytic adaptive tissue reactions is inconsistent (Table 5). ALVAL type reaction has been associated with low wear in several studies, but contradictory results have also been presented. Grammatopoulos et al. stated that the ALVAL reaction correlated moderately positively with increasing wear (Grammatopoulos et al. 2013). Moreover, they also found that hips with minimal wear and pseudotumor had the most severe ALVAL reaction. Our results, which suggest that failed MoM hips pose several different entities that vary in etiology, offer an explanation for the inconsistent findings between the wear and ALVAL responses reported in previous studies. Thus, an ALVAL response or ALVAL-type response may be the result of two different entities that may also differ in wear

characteristics, as suggested in our study. This dualistic nature of ALVAL may explain why previous studies have had conflicting results. We therefore suggest that immunologic/hypersensitivity response to metal wear debris is an important cause of ALVAL-type responses. However, in a subset of patients, there is a higher threshold of wear required before which an abrasion-induced ALVAL-type reaction starts to develop.

A recent study by Ricciardi et al. was very similar to ours (Ricciardi et al. 2016). The major difference was that they also included non-MoM hips and aimed to investigate the subtypes of histopathological findings related to the corrosion products released from a variety of hip replacements. They defined four subtypes based on the available literature. One of the four subtypes was the macrophage-dominated pattern. The second subtype was "mixed lymphocytic and macrophagic with or without features associated with hypersensitivity/allergy or response to particle toxicity." This subtype is equal to the traditional ALVAL-type reaction that we suggested was dualistic in nature. Ricciardi et al. also suggested the division of this subtype based on the presence of hypersensitivity features. The third subtype described by Ricciardi was predominantly sarcoid-like granulomatous response. Our analysis did not, however, suggest this as a separate entity. This could be due to several factors. First, the majority of cases with sarcoid-like pattern were seen in the non-MoM hips with dual modular tapers. We do not suggest, therefore, that the sarcoid-like pattern would constitute a separate entity in MoM hips. Second, the segregation methods used in our study relied on the association of several histopathological variables. Thus, if the sarcoid-like granulomas developed in isolation, it is unlikely that this entity would be identified in the segregation analysis.

## 6.6 Etiopathogenesis of ARMD

The etiopathogenesis of the adverse reactions related to metal debris has been widely researched. However, the mechanisms and development of these often destructive lesions are still poorly understood. Attempts have been made to correlate implant wear, SF metal concentrations, tissue metal concentrations, WB metal ion levels and histopathological findings to better understand the mechanisms (Table 5). Results have been notably discrepant. It has been mathematically proven that most published research findings are false (Ioannidis 2005), and this field of science is likely to be no exception. The replication of

scientific results is considered critical and improves the chance of the findings being true (Moonesinghe et al. 2007). Thus, we aimed to investigate several discrepant associations in the field of ARMD etiopathogenesis. Also, we aimed to create novel approaches for the study of ARMD etiopathogenesis and to investigate new hypotheses to be further tested and validated in future studies.

## 6.6.1 Role of wear

Bearing wear volume correlated weakly with the thickness of macrophage sheets and moderately with the degree of necrosis. Volumetric wear rate correlated moderately with the degree of necrosis, but correlation with macrophage sheet thickness could not be established. Neither wear volume nor volumetric wear rate correlated with lymphocytic cuffing.

Metal debris originating from the bearing surfaces and/or trunnion has been shown to have cytotoxic effects on cells (Catelas et al. 2001, Petit et al. 2004, Scharf et al. 2014). It has been suggested that the cytotoxicity of metal debris further leads to tissue destruction and macrophage recruitment to clear the tissue and metal debris (Mahendra et al. 2009, Ricciardi et al. 2016). In support of this, we observed a correlation between implant wear and the number of macrophages as well as the degree of necrosis but not with the number of lymphocytes. Similar findings were made in a study by Grammatopoulos et al. (Grammatopoulos et al. 2013). Other studies have also noted an association between wear and macrophages but not with necrosis (Campbell et al. 2010, 2018a, Ebramzadeh et al. 2014). Langton et al., however, did not find correlation between wear and the number of macrophages or the amount of necrosis (Langton et al. 2011b). We also observed that the presence of granulomas was associated with increased total wear volume. Granulomas are thought to form in response to a high number of wear particles and our results support this idea (Gallo et al. 2014). Metal hypersensitivity leading to type IV response with strong lymphocytic infiltration has been suggested as a cause of failure in those patients with low wear (Campbell et al. 2010, Ebramzadeh et al. 2011, 2014, Nawabi et al. 2014). Contrary to our hypothesis, this was not observed in our study. In fact, patients with heavy lymphocytic cuffing had a higher wear mean rate than patients with lower numbers of lymphocytes. These findings suggest that excessive metal debris accumulation, not metal hypersensitivity, was the cause of lymphocytic cuffing, at least in some patients with hip resurfacing. Grammatopoulos et al. also found that the presence of

lymphocytes was associated with higher linear wear rate (Grammatopoulos et al. 2013). Further, they reported the presence of a patient subgroup with hypersensitivity-related histopathological findings (lymphocytes and necrosis) and simultaneous low bearing wear, suggesting metal hypersensitivity as a cause of failure in those patients.

## 6.6.2   Role of synovial fluid metal

Synovial fluid cobalt correlated moderately with lymphocyte cuff thickness and necrosis. Chromium correlated only with necrosis. Reito et al. reported similar correlations (Reito et al. 2015b). These results support the hypothesis of metal debris having direct cytotoxic effects on tissues (Mahendra et al. 2009). Interestingly, only cobalt levels correlated with lymphocyte cuff thickness and this correlation was positive. This finding suggests that cobalt in synovial fluid may be dose-dependently relevant in the pathogenesis of lymphocytic ALVAL tissue response, at least in patients with hip resurfacings.

## 6.6.3   Role of whole blood metal ions

Whole blood metal ions have been investigated in relation to ARMD etiopathogenesis with discrepant results. In study I, we noted a correlation between WB metal ion levels and necrosis as well as macrophage sheet thickness. No correlation between WB metal ion levels and lymphocyte cuff thickness was observed. These results support the hypotheses of both foreign-body response and cytotoxic response in response to metal wear debris (Mahendra et al. 2009, Campbell et al. 2010, Grammatopoulos et al. 2013). Paukkeri et al. reported that patients with high numbers of lymphocytes in flow-cytometry had low levels of metal ions in blood compared with patients with a lower number of lymphocytes. They suggested type IV metal hypersensitivity response in patients with low metal ion levels in blood. We did not find a similar association. Other studies have also failed to find a similar association (Lohmann et al. 2013, Grammatopoulos et al. 2017b).

Interestingly, in study II, we did not find similar correlations between WB metal ion levels and histopathological findings. Only WB cobalt and metal particle load within macrophages correlated, but only in the THA group. One major difference between the two studies is that in study I we included only the hip resurfacings of

one manufacturer, whereas in study II several hip replacements (both THAs and hip resurfacings) from different manufacturers were included. As discussed earlier, there are differences in the metallurgy between manufacturers. This may therefore have hindered possible associations. Furthermore, the hip resurfacing group in study II was relatively small and likely underpowered to detect meaningful associations. Besides, in the THA group, metal released from trunnion surfaces likely behaves differently than the metal released from bearing surfaces (Xia et al. 2017).

### 6.6.4 Role of periprosthetic tissue metal

Periprosthetic tissue metal concentrations correlated only with metal particle load within macrophages but not with other histopathological findings. However, tissues with lymphocytic infiltration had lower amounts of chromium compared with tissues with no lymphocytic infiltration. To the best of our knowledge, only one previous study has investigated the periprosthetic metal content in relation to histopathological findings in patients with failed MoM hip arthroplasties (Lohmann et al. 2013). Lohmann et al. found that high periprosthetic tissue metal content (chromium, cobalt and nickel combined and separately) was associated with a lymphocyte-dominated response and low metal content with a macrophage-dominated response. We did not find a similar association – in fact our results favor the opposite. There are some weaknesses in the study by Lohmann et al. which may have distorted the outcome. First, the small number of cases in their study is likely to be a limiting factor. There were only five patients in the macrophage-dominated group and 22 patients in the lymphocyte-dominated group. The high prevalence of the lymphocyte-dominated response compared with the macrophage-dominated response is neither supported by previous studies nor the results of our studies (Campbell et al. 2010, Natu et al. 2012, Grammatopoulos et al. 2013). Furthermore, the mean values for tissue metal concentration were calculated and compared between the two groups. With nonparametric variables, this is not a valid statistical method.

A recent review suggested that periprosthetic tissue metal concentrations may correlate more accurately with the histology than serum metal ion levels (Athanasou 2016). Our results do not support this hypothesis. We found that tissue metal concentrations had poor or non-existent correlations with histological findings. However, tissues with lymphocytic infiltration had lower amounts of

chromium compared with tissues with no lymphocytic infiltration. This finding alone supports the hypothesis of hypersensitivity as a cause of failure in patients with low-wearing MoM hip implants. However, we could not establish correlation between the number of lymphocytes and periprosthetic chromium concentration, which makes it difficult to draw conclusions in light of the overall results.

### 6.6.5    Role of metal debris origin

We found that periprosthetic tissues retrieved from patients with total hip replacements evinced more severe necrosis and more lymphocytes compared with tissues retrieved from patients with hip resurfacings. Taper wear debris has been suggested to be more immunogenic and cytotoxic than bearing wear debris (Langton et al. 2013a, Xia et al. 2017). Xia et al. compared tissues from patients with dual-modular non-MoM implants, MoM THA and MoM hip resurfacings. In dual-modular implants there are two modular junctions which serve as a source of trunnion wear, whereas in THA there is one modular junction and one bearing couple. In hip resurfacing, there are no modular junctions at all, and all wear debris originates from the bearing surfaces. Xia et al. found that tissues from patients with dual-modular implants had the highest amounts of lymphocytes and tissue destruction, whereas tissues from THA patients had lower amounts and, ultimately, tissues from hip resurfacing patients had the lowest amounts. This was despite the fact that tissues from patients with dual-modular non-MoM implants had the de facto lowest amount of metal debris. Also, patients with dual-modular implants had the shortest time to failure. The authors concluded that trunnion wear is likely more immunogenic and cytotoxic than bearing wear debris, leading to rapid failure. Our results support these findings and suggest that taper wear may cause more tissue destruction than bearing wear manifesting as substantially higher failure rates for THAs than hip resurfacings despite similar amounts of metals in the periprosthetic tissues.

### 6.6.6    Role of intrinsic factors

In the present dissertation, we found that there were notable differences in the histological findings between patients revised for ARMD, that is, the between-subject variability was high. Heterogeneity has been characteristic for the results of ARMD research (Campbell et al. 2014). Most importantly, however, we found no

statistically or clinically significant differences in most of the histological and imaging findings between the left and right hips of the same patient, meaning that the within-subject variability in histological and imaging findings was low. Further, the majority of the patients had similar findings in both hips in several key histological variables. This was despite the fact that there was a clinically and statistically significant difference in the amount of wear volume between the sides, that is, there was a difference in the extrinsic factor between the sides. There are no clearly defined boundaries for abnormal versus normal wear, but volumetric wear rates exceeding 1 mm$^3$/year are generally considered abnormal (Sidaginamale et al. 2013, Cook et al. 2019). As the median difference of 15.4 mm$^3$ in wear volume between contralateral sides measured in our study translates into remarkably abnormal yearly volumetric wear rate needed to generate that difference, we thus feel safe to consider the difference in median wear volume between the sides to be clinically significant.

The contribution of host-specific factors in the pathogenesis of ARMD has been suggested in numerous previous studies, manifesting as patient susceptibility of different levels (Mabilleau et al. 2008, Campbell et al. 2010, Donell et al. 2010, Ebramzadeh et al. 2011, Hart et al. 2012b, Matthies et al. 2012, Ebramzadeh et al. 2014, Athanasou 2016). However, to the best of our knowledge, there have been no previous studies that have actually assessed the role of intrinsic factors in the pathogenesis. On the contrary, there have been many studies that have investigated implant wear or the indirect markers of wear and the development of ARMD; however, the results of these studies are very discrepant as discussed in previous chapters. High wear or high blood metal ion levels resulting from high wear are associated with the risk for the development of ARMD (Langton et al. 2010, Hart et al. 2014). However, adverse reactions have been noted in patients with both high and low wearing hip implants (Campbell et al. 2010, Ebramzadeh et al. 2011, Kwon et al. 2011, Langton et al. 2011b, Matthies et al. 2012). In a systematic review, no clear dose-response relationship between wear and ARMD could be established (Campbell et al. 2014). We observed symmetry of histological findings between contralateral hips of the same patients despite differing amounts of wear. In addition, the distribution of wear volume between the sides was similar in patients with symmetrical versus asymmetrical histological and imaging findings. Further, patients with bilateral pseudotumors had similar amounts of wear volume in their hips as patients with no pseudotumor on either side. Our finding suggests that there are intrinsic factors that markedly contribute to the pathogenesis of ARMD that dictate the type of tissue response and the development of

pseudotumors, in addition to extrinsic factors, such as volume and the type of metal wear debris. Further, it is likely that there are differences in these intrinsic factors between patients as some develop aggressive tissue responses despite low-wearing implant, whereas some tolerate large amounts of wear. Various terms, such as patient susceptibility, have been used to describe this phenomenon that clinicians have observed. (Matthies et al. 2012).

A cohort of patients with bilateral MoM hips forms an excellent research framework to investigate and compare the role of intrinsic and extrinsic factors in the pathogenesis. The logic of this reasoning is illustrated in Figure 15. We are aware of only three previous studies that have compared the characteristics of ARMD between the sides in patients with bilateral MoM hip replacements. Madanat et al. compared MRI findings between left and right hips in patients with bilateral MoM hip replacements (Madanat et al. 2015). They found that the soft tissue reaction observed in MRI was symmetrical between the sides in most patients, both in sequentially and simultaneously implanted hips. In support of their findings, we report similar symmetry for the presence of MRI-confirmed pseudotumor between the sides. Another study by Pandit et al. consisted of four revised patients with bilateral MoM hips (Pandit et al. 2008b). All patients had developed a necrotic pseudotumor in both hips. In histopathological analysis, both hips of each patient had similar findings (necrosis, macrophages, lymphocytes). However, no wear data were included in the study and the histology was descriptive, not semiquantitatively scored. A recent study by Uchihara et al. included patients with both unilateral and bilateral MoM hips that had been revised for ARMD (Uchihara et al. 2018). They compared histological findings between left and right hips in the bilateral patients as well as histological findings between unilateral and bilateral patients. In addition, time-to-failure was compared between these two groups. The histological findings (necrosis, macrophages, lymphocytes) between the left and right hips of the bilateral patients were found to be symmetrical in the majority of cases, similar to the findings of the present study. However, we observed that there were differences in the grade of necrosis between the sides, while Uchihara et al. did not semiquantitatively grade necrosis. Further, there were no differences in the histological findings or time-to-failure between unilateral and bilateral patients in their study. Uchihara et al. concluded that the implantation of a MoM hip does not appear to lead to sensitization to metal debris that would in turn lead to poor clinical performance or a different tissue response in the second MoM hip. However, they did not discuss the significance of their findings in the context of intrinsic factors contributing to the similarity of the

tissue response between the contralateral hips in bilateral patients. Furthermore, their sample size was rather small (10 patients) and no wear data of the MoM hips were presented in the study. These three previous studies conducted on bilateral MoM patients are in agreement with our findings and support the hypothesis of an individual host response dictated by intrinsic factors as a significant contributor in the development of soft tissue reactions leading to failure of the hip.
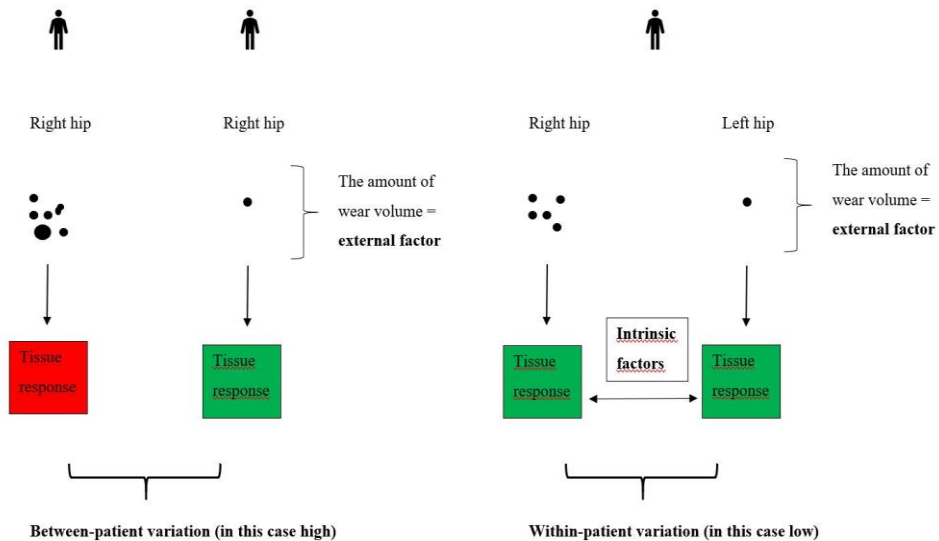


**Figure 15.** Illustration of between- and within-patient variation and the effect of external and intrinsic factors on these.

## 6.6.7 Limitations in our studies

There are several limitations regarding the ability to draw conclusions of ARMD etiopathogenesis from histopathological studies and wear markers. Many of these limitations are common to all of the studies included in this dissertation. First, the histological scoring was performed by one observer only. However, multiple microtome sections of each sample were made and analyzed by a senior musculoskeletal pathologist well acquainted with ARMD histopathology. The samples obtained from soft tissues perioperatively might not represent the overall response of the synovia. Several samples would minimize this variation, but this approach has practical limitations since tissue preservation is important during revision surgery. However, Vaculova et al. found low intrapatient variability in

histological findings between different sampling sites of synovial tissue (Vaculova et al. 2018). Thus, a tissue-preserving approach would seem justified and accurate. Furthermore, there might be variation in cell counts among different sections from the same sample. However, we think that this sampling bias was reduced significantly due to the large group of cases.

The scoring methods we used are semiquantitative and only give indirect information on the underlying pathological processes. Further, the histological findings only reflect one time point, and thus the individual natural history of ARMD cannot be determined. Although we identified four different subtypes of ARMD in Study III, it is possible that some of these reflect the same response at a different time point. In addition, it is possible that each revised hip may evince several different inflammatory responses. Therefore, there is some overlapping and inconsistencies when the distribution of each histological variable among clusters is interpreted. For example, there are cases with a minor grade of necrosis in the "Cytotoxic" group even though we suggested that the hallmark of this group is synovial necrosis. Cluster analysis does not allow overlapping groups. Hence, cases presenting more than one possible different response are forced to one of the remaining clusters. These specific cases cannot be identified during the clustering process, and they are merged to other groups. Since cluster analysis readily revealed different groups, however, we do not consider the overlapping of patterns to have been a significant problem. The prevalence of histological variables may not, however, be representative of the true prevalence since they do not develop in isolation from each other.

Although we found correlation between wear, SF and WB metal ion levels and histopathological findings (Studies I and II), correlation does not necessarily mean full causation. There are many confounding variables. For example, the measurement of WB metal ion levels was not standardized to a certain time interval prior to revision surgery but instead depended on the individual follow-up scheme of the patient. Thus, the WB metal ion levels may not accurately reflect the stages leading to revision surgery. However, most implants are expected to wear at a steady state rate after the initial running-in period (Clarke et al. 2000). Furthermore, WB metal ion levels are likely to not only depend on the amount of metal wear debris produced but also on the capability of macrophage lysosomes to digest the particulate debris into metal ions (Xia et al. 2011). Cellular response of the synovial tissues may affect the way metal debris is handled, and thus affect metal ion levels (Langton et al. 2018).

In Study II, we did not find any evidence that periprosthetic tissue metal concentrations were related to characteristics of tissue responses, but we only measured the gross amounts of metals and not the forms in which the metals were present (ions versus particles, oxidation status, bonding with host proteins, etc.). The amount of metal measured in our study reflects the total amount of all forms combined. The tissue samples used for metal measurement were rather small (approx. 0.3 g) and may not have completely reflected the average metal concentration of the whole synovium. Finally, we were not able to quantify the amount of taper wear debris generated in THAs, which may play a role in the pathogenesis.

## 6.6.8    Discrepancies in the literature

Previous literature on ARMD pathogenesis is vastly discrepant (Table 5). Several possible explanations for this exist. Based on our findings and experience, we suggest the following. First, as we demonstrated in Study III, ARMD is not one or two entities – instead, four different subtypes likely exist. Most previous studies have approached ARMD in dualistic nature – hypersensitivity-related lymphocytic response and wear-related foreign-body response. Second, as has previously been suggested, patients often present a combination of adaptive and innate immunity features in their tissues, and there is overlap between different responses (Berstock et al. 2014). Also, some patients may be susceptible to wear debris, develop hypersensitivity/ALVAL responses and still have poorly performing, high wearing implants. Third, methods between studies are very heterogenous. Implant wear has been measured directly (linear, volumetric, wear rates) and several indirect measures have been used (synovial fluid, tissue concentrations, whole blood). Further, the histological grading of tissues has not been standardized and a range of scoring methods have been used (Campbell et al. 2018b). These heterogeneities make comparison between studies challenging. Moreover, a recent study suggested that current scoring methods show poor intraclass correlation coefficients, and thus might not be reproducible (Smeekes et al. 2017). Histological grading is only semiquantitative and is likely dependent on the observer. Tissue samples only reflect a small portion of the inflamed synovia. Fourth, there is a lack of a clear terminology. Terms such as ARMD, pseudotumor and ALVAL are used interchangeably, although they have different meanings (Athanasou 2016). Fifth, the reported patient populations have mostly been small. Small sample sizes lead to

inadequate statistical power, a higher chance of false positives being reported and low reproducibility of the findings (Button et al. 2013). Furthermore, many studies are performed at tertiary centers where patients are referred from elsewhere. This creates a high possibility of selection bias. In conclusion, there are numerous possible causes for the discrepant findings in the field of ARMD etiopathogenesis. Therefore, the results should be interpreted with caution and replicated.

## 6.6.9    Summary of the etiopathogenesis

We observed four different subtypes of ARMD. The dualistic nature of ALVAL is a novel finding. ARMD can be divided into wear-related foreign-body response, wear-related cytotoxic response, wear-related ALVAL response or hypersensitivity-related ALVAL response. Our study was exploratory in nature, and therefore it should be replicated independently to confirm the results. Wear of the implants was highly variable, and ARMD can also result from low wear. Implant wear correlated with grade of necrosis and macrophages. Although the literature is incongruent, an association between wear and macrophages seems likely when viewed as a whole (Table 5). The association between wear and necrosis is more uncertain as only ours and one previous study have reported a positive association (Grammatopoulos et al. 2013). Necrosis may result from immunological cascades and/or from the direct cytotoxic effects of metal ions, which may create a discrepancy. We did not find correlations between lymphocytic cuff thickness and implant wear or any of the indirect wear measures. Mixed results have been published previously (Table 5). We suggest this is in part due to the dualistic nature of ALVAL we observed – ALVAL may develop due to wear or the susceptibility of the patient or both. Furthermore, the pathogenesis of ARMD in general appears to be affected by individual host-factors that likely leads to different levels of susceptibility to metal debris between patients.

## 6.7    Implications for clinical practice

Patients with MoM implants present a challenge for clinicians. Our results confirm that ARMD may develop both in the presence of a high-wearing, poorly performing implant as well as in a low-wearing implant. High wear is associated with macrophage inflammation and possibly necrosis. WB metal ion levels offer a

reliable measure to estimate the in vivo wear performance of the implant and the chance of adverse soft-tissue reactions. However, aggressive soft-tissue lesions may still be present despite low WB metal ion levels. We have illustrated the meaning of WB metal ion levels in the follow-up of patients in Figure 16. Individual host-factors affect the pathogenesis in addition to external factors, such as wear debris. Bilateral patients define a distinct subset in terms of follow-up. Based on our findings, the development of a pseudotumor on the other side is likely to lead to the development of a similar lesion on the contralateral side.
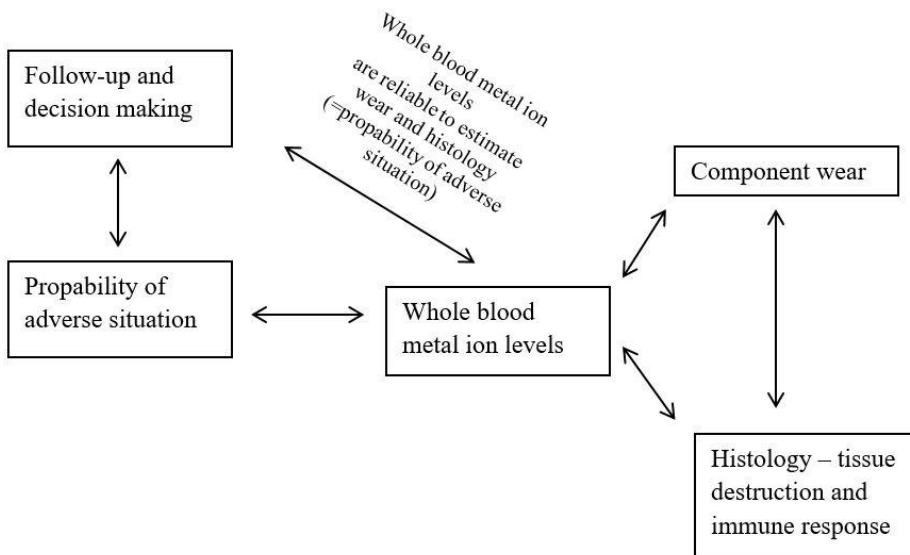


**Figure 16.** A schematic diagram for the relationships between whole blood metal ion levels, wear, and histology of the periprosthetic tissues in the follow-up of patients.

# 7   CONCLUSIONS AND FUTURE PROSPECTS

The primary aim of this dissertation was to study ARMD etiopathogenesis. We analyzed tissue samples histologically to investigate the type of tissue response. We then further assessed possible associations between the tissue responses and different measures of wear (volumetric implant wear, WB and SF metal ion levels, tissue metal concentrations).

Wear volumes ranged from low to manifold higher. The same was true for periprosthetic tissue, WB and SF metal concentrations. WB and SF metal ion levels correlated well with wear volume and volumetric wear rate. Thus, they offer indirect information for surgeons about the wear process of the implant. WB measurement is less invasive, and thus more advisable.

We observed four different histopathological subtypes of ARMD: foreign-body response, cytotoxic response, immunological/hypersensitivity ALVAL and wear-related ALVAL-type response. The dualistic nature of ALVAL response is a novel finding. Our results show that ARMD is not one or two separate entities but four. This further helps to explain many of the discrepancies seen in the previous literature.

Numerous studies have investigated the relationship between implant wear and indirect measures of wear and the histopathological characteristics of the tissue responses. Results have been discrepant. We found correlation between wear and both number of macrophages and grade of necrosis, supporting the hypotheses of foreign-body response and cytotoxic response. Periprosthetic tissue metal concentrations, contrary to what we hypothesized and what has been suggested in the previous literature, did not correlate with histological findings.

Patient susceptibility has been suggested as a key factor in the pathogenesis of ARMD as some patients tolerate high amounts of wear and some develop ARMD in the presence of a low-wearing implant. No studies have been conducted that study the presence of such susceptibility. We found that intrinsic factors determine the type of tissue response in addition to external factors, such as wear. This lends support to the hypothesis of patient susceptibility.

Numerous questions remain unanswered regarding ARMD pathogenesis. Why are some patients more susceptible than others? Can we identify these patients

somehow to better allocate follow-up? Do patients with different types of ARMD have different outcomes of revision surgery? Future studies should try to standardize the histological grading of tissues and reporting to make the results more comparable. Multimodal investigations combining clinical patient data, soft-tissue imaging findings, implant retrieval data, metal ion levels and histopathological findings are needed. Furthermore, flow-cytometry seems a viable, more quantitative and less observer-dependent option to assess the tissue response. Another histological method would be to analyze cellular response based on the cell counts (macrophages, lymphocytes) per high-power fields instead of analyzing the thickness of cellular layers. Immunohistochemical methods could further be combined to deepen the understanding of the roles of macrophages and lymphocytes.

Understanding the differences between taper wear debris and bearing wear debris is of the utmost importance. Although the use of MoM hip replacements has ceased, THAs with other bearing surfaces are still frequently used. The trunnion is still a source for metal debris and ARMD has caused the failure of MoP THAs as well (Whitehouse et al. 2015). Minimizing the potential for these reactions is therefore important.

Finally, although more than a decade of ARMD research has been pursued, the pathogenesis remains only partially understood. A summary of the current knowledge based on ours and previous research is presented in Figure 17. Substantial numbers of patients with MoM replacements remain in follow-up. Mastering the pathogenesis and understanding the differences between patients may be of relevance in clinical decision making – to revise or to continue follow-up? Further, it is important to understand the pathogenesis of ARMD thoroughly in order to design safer hip implants in the future and to avoid the same mistakes that were made with the current generation MoM hip replacements.
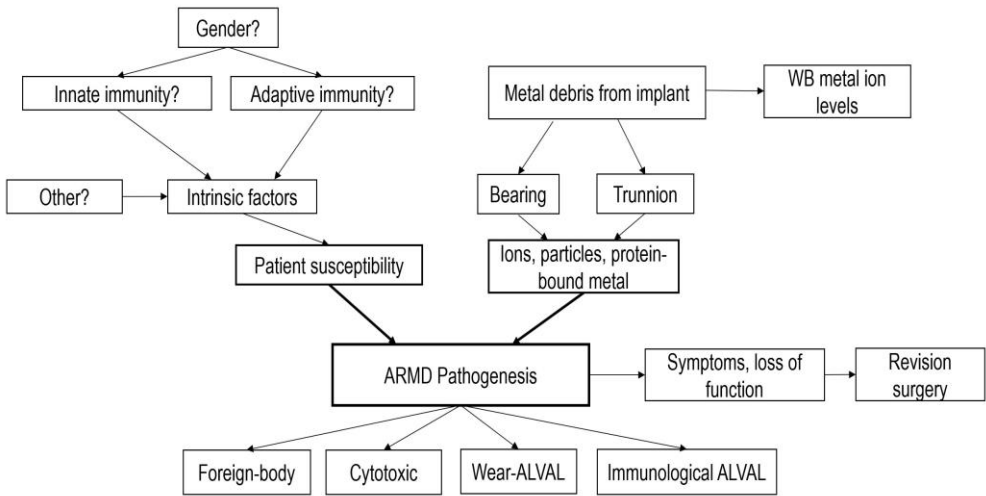
**Figure 17.** A summary of the current knowledge regarding ARMD pathogenesis. Question marks refer to hypotheses proposed in the literature which are yet to be confirmed.

# 8   ACKNOWLEDGEMENTS

This doctoral dissertation was carried out at the Faculty of Medicine and Health Technology of the Tampere University and at Coxa Hospital for Joint Replacement during the years 2014-2020.

First of all, I would like to thank my first supervisor, docent, orthopaedic surgeon and Coxa research director Antti Eskelinen, MD, PhD, for all the effort he has put into this research project. He made it possible for a medical student like me to learn how to do research, become part of the international research community and develop important skills. Further, I greatly appreciated his understanding and support when things did not proceed as planned. Antti has the remarkable gift of being able to view things from a wider perspective in which research is just one part. I also thank him for the memoizable discussions we had about existential matters and the important things in life. Antti's way of emphasizing wellbeing, health and family before other pursuits in life took a great deal of stress from me as a PhD student and helped put things into perspective. As a result, I never felt any pressure from that direction.

I wish to also express my gratitude to my second supervisor, docent and orthopaedic resident Aleksi Reito, MD, PhD. As a young, hardworking and passionate researcher, Aleksi is a trailblazer and has encouraged others by his example. He has taught me the importance of having a critical mindset in the interpretation of research findings. Furthermore, he has highlighted the importance of questioning current truths which cease to be truths when they are challenged by new and better evidence. Also, his intense devotion to medical statistics has sown in me a seed to learn the subject better.

My sincere gratitude also goes to my third supervisor, docent and pathologist Jyrki Parkkinen, MD, PhD.Jyrki's vast knowledge of pathology was integrated with knowledge from the fields of orthopaedic and bioengineering sciences to create a translational research approach. Jyrki has done an enormous job by carefully going through all the pathological samples gathered at Coxa. Moreover, without him I would not have understood even the small part of histopathology I now consider myself to understand.

Tampere, March 2020

Lari Lehtovirta

# 9   REFERENCES

Affatato S, Spinelli M, Zavalloni M, Mazzega-Fabbro C, Viceconti M. Tribology and total hip joint replacement: Current concepts in mechanical simulation. Med Eng Phys. 2008;30(10):1305–17.

Agne MT, Underwood RJ, Kocagoz SB, MacDonald DW, Day JS, Parvizi J, Kraay MJ, Mont MA, Klein GR, Cates HE, Kurtz SM. Is There Material Loss at the Backside Taper in Modular CoCr Acetabular Liners? Clin Orthop Relat Res. Springer US; 2015;473(1):275–85.

Almousa SA, Greidanus N V., Masri BA, Duncan CP, Garbuz DS. The natural history of inflammatory pseudotumors in asymptomatic patients after metal-on-metal hip arthroplasty. Clin Orthop Relat Res. 2013;471(12):3814–21.

Amstutz HC, Le Duff MJ. Background of metal-on-metal resurfacing. Proc Inst Mech Eng Part H J Eng Med. 2006;220(2):85–94.

Amstutz HC, Grigoris P. Metal on metal bearings in hip arthroplasty. Clin Orthop Relat Res. 1996;(329 Suppl):S11-34.

Anderson H, Toms AP, Cahir JG, Goodwin RW, Wimhurst J, Nolan JF. Grading the severity of soft tissue changes associated with metal-on-metal hip replacements: Reliability of an MR grading system. Skeletal Radiol. 2011;40(3):303–7.

Anissian HL, Stark A, Gustafson A, Good V, Clarke IC. Metal-on-metal bearing in hip prosthesis generates 100-fold less wear debris than metal-on-polyethylene. Acta Orthop. Informa UK Ltd; 1999;70(6):578–82.

AOANJRR. The Australian Orthopaedic Association National Joint Replacement Registry. Annual Report. 2007.

AOANJRR. Australian Orthopaedic Association National Joint Replacement Registry. 2017 Annual Report. 2017.

AOANJRR. Australian Orthopaedic Association National Joint Replacement Registry. 2018 Annual Report. 2018.

ASM International. Handbook of Materials for Medical Devices. Davis J, editor. ASM Int. 2003.

Athanasou NA. The pathobiology and pathology of aseptic implant failure. Bone Joint Res. Athanasou; 2016;5(5):162–8.

Back DL, Dalziel R, Young D, Shimmin A. Early results of primary Birmingham hip resurfacings: AN INDEPENDENT PROSPECTIVE STUDY OF THE FIRST 230 HIPS. J Bone Jt Surg - Br Vol. 2005;87-B(3):324–9.

Barceloux DG. Cobalt. J. Toxicol. - Clin. Toxicol. Taylor & Francis; 1999. p. 201–

16.

Becker A, Dirix Y. Wear measurements on retrieved metal-on-metal bearings: a high accuracy 3D measurement approach. Proc 55th Annu Meet Orthop Res Soc. 2009.

Beckstead JW. Using Hierarchical Cluster Analysis in Nursing Research. West J Nurs Res. SAGE Publications; 2002;24(3):307–19.

Berber R, Khoo M, Cook E, Guppy A, Hua J, Miles J, Carrington R, Skinner J, Hart A. Muscle atrophy and metal-on-metal hip implants: a serial MRI study of 74 hips. Acta Orthop. Taylor & Francis; 2015;86(3):351–7.

Berstock JR, Baker RP, Bannister GC, Case CP. Histology of failed metal-on-metal hip arthroplasty; three distinct sub-types. HIP Int. 2014;24(3):243–8.

Bills P, Blunt L, Jiang X. Development of a technique for accurately determining clinical wear in explanted total hip replacements. Wear. 2007;263(7–12):1133–7.

Bills P, Brown L, Jiang X, Blunt L. A metrology solution for the orthopaedic industry. J Phys Conf Ser. 2005;13:316–9.

Bills PJ, Racasan R, Underwood RJ, Cann P, Skinner J, Hart AJ, Jiang X, Blunt L. Volumetric wear assessment of retrieved metal-on-metal hip prostheses and the impact of measurement uncertainty. Wear. 2012;274–275:212–9.

Boardman DR, Middleton FR, Kavanagh TG. A benign psoas mass following metal-on- metal resurfacing of the hip. J Bone Jt Surg [Br]. 2006;88:402–4.

Böhler M, Kanz F, Schwarz B, Steffan I, Walter A, Plenk Jr H, Knahr K. Adverse tissue reactions to wear particles from Co-alloy articulations, increased by alumina-blasting particle contamination from cementless Ti-based total hip implants. J Bone Jt Surg. 2002;84(1):128–36.

Bolland BJRF, Culliford DJ, Langton DJ, Millington JPS, Arden NK, Latham JM, Langton „ D J. High failure rates with a large-diameter hybrid metal-on-metal total hip replacement CLINICAL, RADIOLOGICAL AND RETRIEVAL ANALYSIS. J Bone Jt Surg [Br]. 2011;93(5):608–15.

Bonnini S, Corain L, Marozzi M, Salmaso L. Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R. Wiley; 2014.

Bozic KJ, Kurtz S, Lau E, Ong K, Chiu V, Vail TP, Rubash HE, Berry DJ. The Epidemiology of Bearing Surface Usage in Total Hip Arthroplasty in the United States. J Bone Jt Surgery-American Vol. 2009;91(7):1614–20.

Bradberry SM, Wilkinson JM, Ferner RE. Systemic toxicity related to metal hip prostheses. Clin Toxicol. Taylor & Francis; 2014;52(8):837–47.

Brown GC, Lockshin MD, Salvati EA, Bullough PG. Sensitivity to metal as a possible cause of sterile loosening after cobalt-chromium total hip-replacement arthroplasty. J Bone Joint Surg Am. 1977;59(2):164–8.

Brown M, Wittwer C. Flow Cytometry: Principles and Clinical Applications in Hematology. Clin Chem. 2000;46(8):1221–9.

Brown SA, Zhang K, Merritt K, Payer JH. In vivo transport and excretion of corrosion products from accelerated anodic corrosion of porous coated F75

alloy. J Biomed Mater Res. Wiley Online Library; 1993;27(8):1007–17.

Button KS, Ioannidis JP a, Mokrysz C, Nosek B a, Flint J, Robinson ESJ, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. Nature Publishing Group; 2013;14(5):365–76.

Campbell P, Beaule PE, Ebramzadeh E, LeDuff M, Smet K De, Lu Z, Amstutz HC. The John Charnley Award: A Study of Implant Failure in Metal-on-Metal Surface Arthroplasties. Clin Orthop Relat Res. 2006;453:35–46.

Campbell P, Ebramzadeh E, Nelson S, Takamura K, De Smet K, Amstutz H. Histological Features of Pseudotumor-like Tissues From Metal-on-Metal Hips. Clin Orthop Relat Res. Springer-Verlag; 2010;468(9):2321–7.

Campbell P, McKellop H, Alim R, Mirra J, Nutt S, Dorr L, Amstutz H. Metal-On-Metal Hip Replacements: Wear Performance and Cellular Response To Wear Particles. Cobalt-Base Alloy Biomed Appl. ASTM International; 1999. p. 193–209.

Campbell P, Park S-H, Lu Z, Ebramzadeh E. Tissue features, like wear, are highly variable in revised ASR metal-on-metal total hip arthroplasties. Orthop Proc. Bone & Joint Publishing; 2018a.

Campbell PA, Kung MS, Hsu AR, Jacobs JJ. Do retrieval analysis and blood metal measurements contribute to our understanding of adverse local tissue reactions? Clin Orthop Relat Res. Springer; 2014;472(12):3718–27.

Campbell PA, Nguyen M, Yuan N. The Use of Semiquantitative Histology for Reporting Periprosthetic Tissue Features: A Review. Beyond Implant Retr Anal Methods Implant Surveill. 2018b. p. 203–14.

Carli, Reuven, Zukor, Antoniou. Adverse Soft-Tissue Reactions Around Non-Metal-on-Metal Total Hip Arthroplasty - A Systematic Review of the Literature. S47 Bull NYU Hosp Jt Dis. 2011;69:47–51.

Carrothers AD, Gilbert RE, Jaiswal A, Richardson JB. Birmingham hip resurfacing: The prevalence of failure. J Bone Jt Surg - Br Vol. 2010;92-B(10):1344–50.

Catelas I, Bobyn JD, Medley JB, Krygier JJ, Zukor DJ, Huk OL. Size, shape, and composition of wear particles from metal–metal hip simulator testing: effects of alloy and number of loading cycles. J Biomed Mater Res Part A. Wiley Online Library; 2003;67(1):312–27.

Catelas I, Campbell PA, Bobyn JD, Medley JB, Huk OL. Wear particles from metal-on-metal total hip replacements: effects of implant design and implantation time. Proc Inst Mech Eng H, J Eng Med. 2006;220(2):195–208.

Catelas I, Petit A, Vali H, Fragiskatos C, Meilleur R, Zukor DJ, Antoniou J, Huk OL. Quantitative analysis of macrophage apoptosis vs. necrosis induced by cobalt and chromium ions in vitro. Biomaterials. Elsevier; 2005;26(15):2441–53.

Catelas I, Petit A, Zukor DJ, Huk OL. Cytotoxic and apoptotic effects of cobalt and chromium ions on J774 macrophages–Implication of caspase-3 in the apoptotic pathway. J Mater Sci Mater Med. Springer; 2001;12(10–12):949–53.

Catelas I, Wimmer MA, Utzschneider S. Polyethylene and metal wear particles: characteristics and biological effects. Semin Immunopathol. Springer; 2011. p. 257–71.

Clarke IC, Good V, Williams P, Schroeder D, Anissian L, Stark A, Oonishi H, Schuldies J, Gustafson G. Ultra-low wear rates for rigid-on-rigid bearings in total hip replacements. Proc Inst Mech Eng H, J Eng Med. 2000;214(4):331–47.

Cohen D. How safe are metal-on-metal hip implants? Br Med J. 2012;344(7846):e1410.

Cook RB, Pearce AR, Culliford DJ, Briant-Evans T, Griffiths JT, Britton JM, Stranks GJ. Retrieval Analysis And Principal Variable Analysis Of 127 M2a-38 Mm™ Metal On Metal Hip Replacements. Biotribology. 2019;19:100102.

Cooper HJ, Urban RM, Wixson RL, Meneghini RM, Jacobs JJ. Adverse Local Tissue Reaction Arising from Corrosion at the Femoral Neck-Body Junction in a Dual-Taper Stem with a Cobalt-Chromium Modular Neck. J Bone Jt Surg. 2013;95(10):865–72.

Cooper HJ, Della Valle CJ, Berger RA, Tetreault M, Paprosky WG, Sporer SM, Jacobs JJ. Corrosion at the head-neck taper as a cause for adverse local tissue reactions after total hip arthroplasty. J Bone Jt Surg - Am Vol. 2012;94(18):1655–61.

Daniel J, Pynsent PB, McMinn DJW. Metal-on-metal resurfacing of the hip in patients under the age of 55 years with osteoarthritis. J Bone Jt Surg. 2004;86(2):177–84.

Davda K, Lali F V, Sampson B, Skinner JA, Hart AJ. An analysis of metal ion levels in the joint fluid of symptomatic patients with metal-on-metal hip replacements. J Bone Jt Surgery, Br Vol. 2011;93(6):738–45.

Davies a P, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. J Bone Jt Surg - Am Vol. 2005;87(1):18–27.

Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Jt Surg - Ser B. 1996;78(2):185–90.

DePuy Orthopaedics. Urgent field safety notice. DePuy ASR™ Articular Surface Replacement and ASR™ XL Acetabular System Recall. August 24, 2010. 2010.

Donell ST, Darrah C, Nolan JF, Wimhurst J, Toms A, Barker THW, Case CP, Tucker JK. Early failure of the Ultima metal-on-metal total hip replacement in the presence of normal plain radiographs. 2010;(11):1501–8.

Doorn PF, McKellop HA, Benya PD, Campbell PA, Worrall J, Amstutz HC. Metal wear particle characterization from metal on metal total hip replacements: Transmission electron microscopy study of periprosthetic tissues and isolated particles. J Biomed Mater Res. 1998;42(1):103–11.

Doorn PF, Mirra JM, Campbell PA, Amstutz HC. Tissue reaction to metal on

metal total hip prostheses. Clin Orthop Relat Res. LWW; 1996;329:S187–205.

Dowson D. Tribological principles in metal-on-metal hip joint design. Proc Inst Mech Eng Part H J Eng Med. 2006;220(2):161–71.

Ebramzadeh E, Campbell P, Tan TL, Nelson SD, Sangiorgio SN. Can Wear Explain the Histological Variation Around Metal-on-metal Total Hips? Clin Orthop Relat Res. Springer US; 2014;473(2):487–94.

Ebramzadeh E, Campbell PA, Takamura KM, Lu Z, Sangiorgio SN, Kalma JJ, De Smet KA, Amstutz HC. Failure modes of 433 metal-on-metal hip implants: how, why, and wear. Orthop Clin North Am. Elsevier; 2011;42(2):241–50.

Eltit F, Wang Q, Wang R. Mechanisms of Adverse Local Tissue Reactions to Hip Implants. Front Bioeng Biotechnol. Frontiers Media SA; 2019;7:176.

Evans EM, Freeman MA, Miller AJ, Vernon-Roberts B. Metal sensitivity as a cause of bone necrosis and loosening of the prosthesis in total joint replacement. J Bone Jt Surg Br. 1974;56-B(4):626–42.

FDA. Class 2 Device Recall Durom cup. 2008.

FDA. Class 2 Device Recall  ABG II Modular Hip Stem. 2012a.

FDA. Class 2 Device Recall Rejuvenate Modular Stems. 2012b.

FDA. US Food & Drug Administration. Metal-on-Metal Hip Implants - Information for Orthopaedic Surgeons. 2013.

FDA. Class 2 Device Recall  Modular REDAPT(TM) Hip Systems. 2016a.

FDA. Class 2 Device Recall  Modular SMF(TM). 2016b.

Fehring TK, Odum S, Sproul R, Weathersbee J. High frequency of adverse local tissue reactions in asymptomatic patients with metal-on-metal THA. Clin Orthop Relat Res. Springer New York LLC; 2014;472(2):517–22.

Firkins PJ, Tipper JL, Saadatzadeh MR, Ingham E, Stone MH, Farrar R, Fisher J. Quantitative analysis of wear and wear debris from metal-on-metal hip prostheses tested in a physiological hip joint simulator. Biomed Mater Eng. Pergamon Press; 2001;11(2):143–57.

Fraga CG. Relevance, essentiality and toxicity of trace elements in human health. Mol Aspects Med. 2005;26(4–5):235–44.

Fujiwara N, Kobayashi K. Macrophages in inflammation. Curr. Drug Targets Inflamm. Allergy. 2005. p. 281–6.

Gallo J, Vaculova J, Goodman SB, Konttinen YT, Thyssen JP. Contributions of human tissue analysis to understanding the mechanisms of loosening and osteolysis in total hip replacement. Acta Biomater. NIH Public Access; 2014;10(6):2354–66.

Gascoyne TC, Dyrkacz RM, Turgeon TR, Burnell CD, Wyss UP, Brandt J-M. Corrosion on the Acetabular Liner Taper from Retrieved Modular Metal-on-Metal Total Hip Replacements. J Arthroplasty. 2014;29(10):2049–52.

Gill IP, Webb J, Sloan K, Beaver RJ. Corrosion at the neck-stem junction as a cause of metal ion release and pseudotumour formation. J Bone Joint Surg Br. 2012;94(7):895–900.

Glyn-Jones S, Pandit H, Kwon YM, Doll H, Gill HS, Murray DW. Risk factors for

inflammatory pseudotumour formation following hip resurfacing. J Bone Jt Surg. 2009;91-B(12):1566–74.

Glyn-Jones S, Roques A, Taylor A, Kwon Y-M, McLardy-Smith P, Gill HS, Walter W, Tuke M, Murray D. The In Vivo Linear and Volumetric Wear of Hip Resurfacing Implants Revised for Pseudotumor. J Bone Jt Surgery-American Vol. 2011;93(23):2180–8.

Goldberg JR, Gilbert JL, Jacobs JJ, Bauer TW, Paprosky W, Leurgans S. A multicenter retrieval study of the taper interfaces of modular hip prostheses. Clin Orthop Relat Res. 2002;(401):149–61.

Goldsmith AAJ, Dowson D, Isaac GH, Lancaster JG. A comparative joint simulator study of the wear of metal-on-metal and alternative material combinations in hip replacements. Proc Inst Mech Eng Part H J Eng Med. SAGE Publications; 2000;214(1):39–47.

Gomez PF, Morcuende JA. Early attempts at hip arthroplasty--1700s to 1950s. Iowa Orthop J. 2005;25:25–9.

Gower JC. A General Coefficient of Similarity and Some of Its Properties. Biometrics. International Biometric Society; 1971;27(4):857.

Grammatopolous G, Pandit H, Kwon Y-M, Gundle R, McLardy-Smith P, Beard DJ, Murray DW, Gill HS. Hip resurfacings revised for inflammatory pseudotumour have a poor outcome. J Bone Joint Surg Br. 2009;91-B(8):1019–24.

Grammatopoulos G, Munemoto M, Pollalis A, Athanasou NA. Correlation of serum metal ion levels with pathological changes of ARMD in failed metal-on-metal-hip-resurfacing arthroplasties. Arch Orthop Trauma Surg. 2017a;137(8):1129–37.

Grammatopoulos G, Munemoto M, Pollalis A, Athanasou NA. Correlation of serum metal ion levels with pathological changes of ARMD in failed metal-on-metal-hip-resurfacing arthroplasties. Arch Orthop Trauma Surg. 2017b;137(8):1129–37.

Grammatopoulos G, Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gill HS, Murray DW. Optimal acetabular orientation for hip resurfacing. J Bone Jt Surg - Br Vol. 2010;92-B(8):1072–8.

Grammatopoulos G, Pandit H, Orth F, Kamali A, Maggiani F, Athanasou N, Glyn-Jones S, Gill HS, Murray DW, Athanasou N, G. G, H. P, A. K, F. M, S. G-J, H.S. G, D.W. M, N. A. The Correlation of Wear with Histological Features After Failed Hip Resurfacing Arthroplasty. J Bone Jt Surg - Ser A. 2013;81(12):1–10.

Gruber FW, Böck A, Trattnig S, Lintner F, Ritschl P. Cystic Lesion of the Groin due to Metallosis: A Rare Long-Term Complication of Metal-on-Metal Total Hip Arthroplasty. J Arthroplasty. Churchill Livingstone; 2007;22(6):923–7.

Guo Z, Sadler PJ. Metals in Medicine. Angew Chemie Int Ed. WILEY-VCH Verlag GmbH; 1999;38(11):1512–31.

De Haan R, Campbell PA, Su EP, De Smet KA. Revision of metal-on-metal

resurfacing arthroplasty of the hip: The influence of malpositioning of the components. J Bone Jt Surg - Br Vol. 2008;90-B(9):1158–63.

Haddad FS, Thakrar RR, Hart AJ, Skinner JA, Nargol AVF, Nolan JF, Gill HS, Murray DW, Blom AW, Case CP. Metal-on-metal bearings: THE EVIDENCE SO FAR. Bone Joint J. 2011;93-B(5):572–9.

Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am. 1969;51(4):737–55.

Harris WH. Osteolysis and particle disease in hip replacement: A review. Acta Orthop. 1994;65(1):113–23.

Hart AJ, Allwood SM, Porter M, Matthies A, Maggiore P, Underwood R, Cann P, Cobb J, Skinner J a. Which Factors Determine the Wear Rate of Large-Diameter Metal-on-Metal Hip Replacements? J Bone Joint Surg Am. 2013;95(8):678–85.

Hart AJ, Ilo K, Underwood R, Cann P, Henckel J, Lewis A, Cobb J, Skinner J. The relationship between the angle of version and rate of wear of retrieved metal-on-metal resurfacings. J Bone Joint Surg Br. 2011a;93-B(3):315–20.

Hart AJ, Matthies A, Henckel J, Ilo K, Skinner J, Noble PC. Understanding Why Metal-on-Metal Hip Arthroplasties Fail. J Bone Jt Surgery-American Vol. 2012a;94(4):e22-1–10.

Hart AJ, Matthies A, Henckel J, Ilo K, Skinner J, Noble PC. Understanding Why Metal-on-Metal Hip Arthroplasties Fail A Comparison Between Patients with Well-Functioning and Revised Birmingham Hip Resurfacing Arthroplasties AAOS Exhibit Selection. J Bone Jt Surgery-American Vol. 2012b;94(4):e22(1-10).

Hart AJ, Matthies AK, Racasan R, Bills PJ, Panagiotidou A, Blunt L, Blunn G, Skinner J. Taper wear contributes only a third of the total volumetirc material loss in large head metal on metal hip replacement. Orthop Proc. 2012c.

Hart AJ, Quinn PD, Sampson B, Sandison A, Atkinson KD, Skinner JA, Powell JJ, Mosselmans JFW. The chemical form of metallic debris in tissues surrounding metal-on-metal hips with unexplained failure. Acta Biomater. Elsevier; 2010;6(11):4439–46.

Hart AJ, Sabah SA, Bandi AS, Maggiore P, Tarassoli P, Sampson B, A. Skinner J. Sensitivity and specificity of blood cobalt and chromium metal ions for predicting failure of metal-on-metal hip replacement. Bone Joint J. 2011b;93-B(10):1308–13.

Hart AJ, Sabah SA, Henckel J, Lloyd G, Skinner JA. Lessons learnt from metal-on-metal hip arthroplasties will lead to safer innovation for all medical devices. Hip Int. 2015;25(4):347–54.

Hart AJ, Sabah SA, Sampson B, Skinner JA, Powell JJ, Palla L, Pajamaki KJ, Puolakka T, Reito A, Eskelinen A, Pajamäki KJJ, Puolakka T, Reito A, Eskelinen A. Surveillance of Patients with Metal-on-Metal Hip Resurfacing and Total Hip Prostheses: A Prospective Cohort Study to Investigate the

Relationship Between Blood Metal Ion Levels and Implant Failure. J Bone Jt Surgery-American Vol. 2014;96(13):1091–9.

Hart, Satchithananda K, Liddle AD, Sabah SA, Mcrobbie D, Henckel J, Cobb JP, Skinner JA, Mitchell AW, Orth F, Mitchell AW. Pseudotumors in Association with Well-Functioning Metal-on-Metal Hip Prostheses. J Bone Jt Surg Am. 2012d;94(4):317–25.

Hasegawa M, Iino T, Sudo A. Immune response in adverse reactions to metal debris following metal-on-metal total hip arthroplasty. BMC Musculoskelet Disord. BioMed Central; 2016;17(1):221.

Health Canada. Health Canada. Metal-on-Metal Hip Implants - Information for Orthopaedic Surgeons Regarding Patient Management Following Surgery. 2012.

Heisel C, Kleinhans JA, Menge M, Kretzer JP. Ten different hip resurfacing systems: Biomechanical analysis of design and material properties. Int Orthop. 2009;33(4):939–43.

Higgs GB, Hanzlik JA, MacDonald DW, Gilbert JL, Rimnac CM, Kurtz SM. Is Increased Modularity Associated With Increased Fretting and Corrosion Damage in Metal-On-Metal Total Hip Arthroplasty Devices?: A Retrieval Study. J Arthroplasty. Churchill Livingstone; 2013;28(8):2–6.

Hothi HS, Berber R, Panagiotopoulos AC, Whittaker RK, Rhead C, Skinner JA, Hart AJ. Clinical significance of corrosion of cemented femoral stems in metal-on-metal hips: a retrieval study. Int Orthop. Springer Berlin Heidelberg; 2016a;40(11):2247–54.

Hothi HS, Berber R, Whittaker RK, Blunn GW, Skinner JA, Hart AJ. The Relationship Between Cobalt/Chromium Ratios and the High Prevalence of Head-Stem Junction Corrosion in Metal-on-Metal Total Hip Arthroplasty. J Arthroplasty. Churchill Livingstone; 2016b;31(5):1123–7.

Hothi HS, Ilo K, Whittaker RK, Eskelinen A, Skinner JA, Hart AJ. Corrosion of Metal Modular Cup Liners. J Arthroplasty. 2015;30:1652–6.

Hothi HS, Matthies AK, Berber R, Whittaker RK, Skinner JA, Hart AJ. The reliability of a scoring system for corrosion and fretting, and its relationship to material loss of tapered, modular junctions of retrieved hip implants. J Arthroplasty. 2014;29(6).

Hothi HS, Panagiotopoulos AC, Whittaker RK, Bills PJ, McMillan RA, Skinner JA, Hart AJ. Damage Patterns at the Head-Stem Taper Junction Helps Understand the Mechanisms of Material Loss. J Arthroplasty. 2017;32(1):291–5.

Huber M, Reinisch G, Nter Trettenhahn G, Zweymüller K, Lintner F. Presence of corrosion products and hypersensitivity-associated reactions in periprosthetic tissue after aseptic loosening of total hip replacements with metal bearing surfaces. Acta Biomater. 2008;5:172–80.

Ilchmann T, Reimold M, Müller-Schauenburg W. Estimation of the wear volume after total hip replacement. Med Eng Phys. 2008;30(3):373–9.

Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. Public Library of Science; 2005;2(8):e124.

Jacobs JJ, Skipor AK, Patterson LM, Hallab NJ, Paprosky WG, Black J. Metal Release in Patients Who Have Had a Primary Total Hip Arthroplasty. J Bone Jt Surg. 1998;80(10):1447–58.

Jacobs JJ, Wimmer MA. An important contribution to our understanding of the performance of the current generation of metal-on-metal hip replacements. J Bone Joint Surg Am. 2013;95(8).

Koper MC, Mathijssen NMC, Witt F, Morlock MM, Vehmeijer SBW. Clinical and Wear Analyses of 9 Large Metal-on-Metal Total Hip Prostheses. PLoS One. Public Library of Science; 2016;11(10):e0163438.

Krenn V, Morawietz L, Perino G, Kienapfel H, Ascherl R, Hassenpflug GJ, Thomsen M, Thomas P, Huber M, Kendoff D, Baumhoer D, Krukemeyer MG, Natu S, Boettner F, Zustin J, Kölbel B, Rüther W, Kretzer JP, Tiemann A, Trampuz A, Frommelt L, Tichilow R, Söder S, Müller S, Parvizi J, Illgner U, Gehrke T. Revised histopathological consensus classification of joint implant related pathology. Pathol - Res Pract. 2014;210(12):779–86.

Kretzer JP, Kleinhans JA, Jakubowitz E, Thomsen M, Heisel C. A meta-analysis of design- and manufacturing-related parameters influencing the wear behavior of metal-on-metal hip joint replacements. J Orthop Res. 2009;27(11):1473–80.

Krishnan H, Krishnan SP, Blunn G, Skinner JA, Hart AJ. Modular neck femoral stems. Bone Joint J. 2013;95-B(8):1011–21.

Kruger J. Fundamental Aspects of the Corrosion of Metallic Implants. In: B. Syrett and A. Acharya, editor. Corros Degrad Implant Mater. ASTM International; 1979. p. 107–27.

Kuba M, Gallo J, Pluháček T, Hobza M, Milde D. Content of distinct metals in periprosthetic tissues and pseudosynovial joint fluid in patients with total joint arthroplasty. J Biomed Mater Res Part B Appl Biomater. John Wiley & Sons, Ltd; 2019;107(2):454–62.

Kwon Y-M, Glyn-Jones S, Simpson DJ, Kamali A, McLardy-Smith P, Gill HS, Murray DW. Analysis of wear of retrieved metal-on-metal hip resurfacing implants revised due to pseudotumours. J Bone Jt Surgery, Br Vol. 2010;92(3):356–61.

Kwon Y-M, Ostlere SJ, McLardy-Smith P, Athanasou NA, Gill HS, Murray DW. "Asymptomatic" pseudotumors after metal-on-metal hip resurfacing arthroplasty: prevalence and metal ion study. J Arthroplasty. Elsevier; 2011;26(4):511–8.

Kwon Y-M, Xia Z, Glyn-Jones S, Beard D, Gill HS, Murray DW. Dose-dependent cytotoxicity of clinically relevant cobalt nanoparticles and ions on macrophages *in vitro*. Biomed Mater. 2009;4(2):025018.

Lainiala O, Eskelinen A, Elo P, Puolakka T, Korhonen J, Moilanen T. Adverse reaction to metal debris is more common in patients following MoM total hip

replacement with a 36 mm femoral head than previously thought: Results from a modern MoM follow-up programme. Bone Jt J. 2014;96B(12):1610–7.

Lainiala O, Reito A, Elo P, Pajamäki J, Puolakka T, Eskelinen A. Revision of Metal-on-metal Hip Prostheses Results in Marked Reduction of Blood Cobalt and Chromium Ion Concentrations. Clin Orthop Relat Res. Springer; 2015;1–9.

Lainiala OS, Moilanen TPS, Hart AJ, Huhtala HSA, Sabah SA, Eskelinen AP. Higher Blood Cobalt and Chromium Levels in Patients With Unilateral Metal-on-Metal Total Hip Arthroplasties Compared to Hip Resurfacings. J Arthroplasty. Elsevier; 2016;31(6):1261–6.

Lainiala OS, Reito AP, Nieminen JJ, Eskelinen AP. Declining Revision Burden of Metal-on-Metal Hip Arthroplasties. J Arthroplasty. 2019;34(9):2058-2064.e1.

Langton DJ, Jameson SS, Joyce TJ, Gandhi JN, Sidaginamale R, Mereddy P, Lord J, Nargol A V. Accelerating failure rate of the ASR total hip replacement. J bone Jt surgeryBritish Vol. 2011a;93(8):1011–6.

Langton DJ, Jameson SS, Joyce TJ, Hallab NJ, Natu S, Nargol AVF. Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: A consequence of excess wear. J Bone Jt Surgery, Br Vol. 2010;92(1):38–46.

Langton DJ, Joyce TJ, Jameson SS, Lord J, Van Orsouw M, Holland JP, Nargol AVF, De Smet KA. Adverse reaction to metal debris following hip resurfacing the influence of component type, orientation and volumetric wear. Bone Joint J. 2011b;93-B(2):164–71.

Langton DJ, Sidaginamale R, Lord J, Joyce T, Natu S, Nargol A. Metal debris release from taper junctions appears to have a greater clinical impact than debris released from metal on metal bearing surfaces. Bone Jt J. 2013a;95(SUPP 1):1.

Langton DJ, Sidaginamale R, Lord JK, Nargol AVF, Joyce TJ. Taper junction failure in large-diameter metal-on-metal bearings. Bone Joint Res. 2012;1(4):56–63.

Langton DJ, Sidaginamale RP, Avery P, Waller S, Tank G, Lord J, Joyce T, Cooke N, Logishetty R, Nargol AVF. Retrospective cohort study of the performance of the pinnacle metal on metal (MoM) total hip replacement: A single-centre investigation in combination with the findings of a national retrieval centre. BMJ Open. 2016;6(4):e007847.

Langton DJ, Sidaginamale RP, Joyce TJ, Bowsher JG, Holland JP, Deehan D, Nargol AVF, Natu S. Aseptic lymphocyte-dominated vasculitis-associated lesions are related to changes in metal ion handling in the joint capsules of metal-on-metal hip arthroplasties. Bone Joint Res. 2018;7(6):388–96.

Langton DJ, Sidaginamale RP, Joyce TJ, Natu S, Blain P, Jefferson RD, Rushton S, Nargol AVF. The clinical implications of elevated blood metal ion concentrations in asymptomatic patients with MoM hip resurfacings: A cohort study. BMJ Open. British Medical Journal Publishing Group;

2013b;3(3):e001541.

Di Laura A, Hothi HS, Henckel J, Kwon Y-M, Skinner JA, Hart AJ. Retrieval Findings of Recalled Dual-Taper Hips. J Bone Joint Surg Am. Wolters Kluwer Health; 2018;100(19):1661–72.

Di Laura A, Quinn PD, Panagiotopoulou VC, Hothi HS, Henckel J, Powell JJ, Berisha F, Amary F, Mosselmans JFW, Skinner JA, Hart AJ. The Chemical Form of Metal Species Released from Corroded Taper Junctions of Hip Implants: Synchrotron Analysis of Patient Tissue. Sci Rep. Nature Publishing Group; 2017;7(1):10952.

Leslie I, Williams S, Brown C, Isaac G, Jin Z, Ingham E, Fisher J. Effect of bearing size on the long-term wear, wear debris, and ion levels of large diameter metal-on-metal hip replacements - An in vitro study. J Biomed Mater Res - Part B Appl Biomater. 2008;87(1):163–72.

Lewinnek GE, Lewis JL, Tarr R, Compere CL, Zimmerman JR. Dislocations after total hip-replacement arthroplasties. J Bone Joint Surg Am. 1978;60(2):217–20.

Liao Y, Hoffman E, Wimmer M, Fischer A, Jacobs J, Marks L. CoCrMo metal-on-metal hip replacements. Phys Chem Chem Phys. NIH Public Access; 2013;15(3):746–56.

Liow MHL, Urish KL, Preffer FI, Nielson GP, Kwon Y-M. Metal ion levels are not correlated with histopathology of adverse local tissue reactions in taper corrosion of total hip arthroplasty. J Arthroplasty. Churchill Livingstone; 2016;31(8):1797–802.

Lohmann CH, Meyer H, Nuechtern J V, Singh G, Junk-Jantsch S, Schmotzer H, Morlock MM, Pflüger G. Periprosthetic tissue metal content but not serum metal content predicts the type of tissue response in failed small-diameter metal-on-metal total hip arthroplasties. J Bone Jt Surg. 2013;95(17):1561–8.

Lombardi Jr A V, Barrack RL, Berend KR, Cuckler JM, Jacobs JJ, Mont MA, Schmalzried TP. The Hip Society: algorithmic approach to diagnosis and management of metal-on-metal arthroplasty. J Bone Joint Surg Br. 2012;94(11 Suppl A):14–8.

Lord JK, Langton DJ, Nargol AVF, Joyce TJ. Volumetric wear assessment of failed metal-on-metal hip resurfacing prostheses. Wear. Elsevier; 2011;272(1):79–87.

Lu Z, Ebramzadeh E. Origins of material loss in highly worn acetabular cups of metal-on-metal total hip replacements. J Orthop Res. John Wiley & Sons, Ltd; 2019;37(1):143–50.

Mabilleau G, Kwon Y-M, Pandit H, Murray DW, Sabokbar A. Metal-on-metal hip resurfacing arthroplasty: A review of periprosthetic biological reactions. Acta Orthop. Taylor & Francis; 2008;79(6):734–47.

Madanat R, Hussey DK, Donahue GS, Potter HG, Wallace R, Bragdon CR, Muratoglu OK, Malchau H. The Symmetry of Adverse Local Tissue Reactions in Patients with Bilateral Simultaneous and Sequential ASR Hip Replacement. J Arthroplasty. Elsevier B.V.; 2015;30(10):1794–8.

Mahendra G, Pandit H, Kliskey K, Murray D, Gill HS, Athanasou N. Necrotic and inflammatory changes in metal-on-metal resurfacing hip arthroplasties: relation to implant failure and pseudotumor formation. Acta Orthop. 2009;80(6):653–9.

Matharu GS. The investigation and management of metal-on-metal hip arthroplasty patients. University of Oxford; 2017.

Matharu GS, Berryman F, Dunlop DJ, Judge A, Murray DW, Pandit HG. Has the threshold for revision surgery for adverse reactions to metal debris changed in metal-on-metal hip arthroplasty patients? A cohort study of 239 patients using an adapted risk-stratification algorithm. Acta Orthop. 2019;90(6):530–6.

Matharu GS, Eskelinen A, Judge A, Pandit HG, Murray DW. Revision surgery of metal-on-metal hip arthroplasties for adverse reactions to metal debris. Acta Orthop. 2018a;89(3):1–11.

Matharu GS, Judge A, Eskelinen A, Murray DW, Pandit HG. What is appropriate surveillance for metal-on-metal hip arthroplasty patients? Acta Orthop. Taylor & Francis; 2018b;89(1):29–39.

Matharu GS, Judge A, Murray DW, Pandit HG. Prevalence of and Risk Factors for Hip Resurfacing Revision. J Bone Jt Surg. 2016;98(17):1444–52.

Matharu GS, Judge A, Pandit HG, Murray DW. Follow-up for patients with metal-on-metal hip replacements: are the new MHRA recommendations justified? BMJ. British Medical Journal Publishing Group; 2018c;360:k566.

Matharu GS, Mellon SJ, Murray DW, Pandit HG. Follow-up of metal-on-metal hip arthroplasty patients is currently not evidence based or cost effective. J Arthroplasty. Elsevier; 2015;30(8):1317–23.

Matharu GS, Nandra RS, Berryman F, Judge A, Pynsent PB, Dunlop DJ. Risk factors for failure of the 36 mm metal-on-metal Pinnacle total hip arthroplasty system. Bone Joint J. 2017;99-B(5):592–600.

Matthies A, Underwood R, Cann P, Ilo K, Nawaz Z, Skinner J, Hart AJ. Retrieval analysis of 240 metal-on-metal hip components, comparing modular total hip replacement with hip resurfacing. J Bone Jt Surg Br. 2011;93(3):307–14.

Matthies AK, Henckel J, Cro S, Suarez A, Noble PC, Skinner J, Hart AJ. Predicting wear and blood metal ion levels in metal-on-metal hip resurfacing. J Orthop Res. 2013a;32(1):167–74.

Matthies AK, Skinner JA, Osmani H, Henckel J, Hart AJ. Pseudotumors are common in well-positioned low-wearing metal-on-metal hips. Clin Orthop Relat Res. Springer; 2012;470(7):1895–906.

Matthies, Racasan R, Bills P, Blunt L, Cro S, Panagiotidou A, Blunn G, Skinner J, Hart AJ. Material loss at the taper junction of retrieved large head metal-on-metal total hip replacements. J Orthop Res. Wiley Online Library; 2013b;31(11):1677–85.

McKellop H, Park S-H, Chiesa R, Doorn P, Lu B, Normand P, Grigoris P, Amstutz H. In Vivo Wear of 3 Types of Metal on Metal Hip Prostheses During 2 Decades of Use. Clin Orthop Relat Res. 1996;329:S128–40.

McKellop HA, Hart A, Park SH, Hothi H, Campbell P, Skinner JA. A lexicon for wear of metal-on-metal hip prostheses. J Orthop Res. 2014;32(9):1221–33.

McMinn, Treacy, Lin P. Metal on Metal Surface Replacement of the Hip: Experience of the McMinn prosthesis. Curr Orthop Pract. 1996;329:S89–98.

McMinn DJW. Development of metal/metal hip resurfacing. HIP Int. 2003. p. S41–53.

Mellon SJ, Grammatopoulos G, Andersen MS, Pegg EC, Pandit HG, Murray DW, Gill HS. Individual motion patterns during gait and sit-to-stand contribute to edge-loading risk in metal-on-metal hip resurfacing. Proc Inst Mech Eng Part H J Eng Med. 2013a;227(7):799–810.

Mellon SJ, Liddle AD, Pandit H. Hip replacement: Landmark surgery in modern medical history. Maturitas. 2013b;75(3):221–6.

Merritt K, Crowe TD, Brown SA. Elimination of nickel, cobalt, and chromium following repeated injections of high dose metal salts. J Biomed Mater Res. Wiley Online Library; 1989;23(8):845–62.

MHRA. Medicines and Healthcare Products Regulatory Agency. Medical device alert: All metal-on-metal (MoM) hip replacements (MDA/2010/033). 2010.

MHRA. Medicines and Healthcare products Regulatory Agency (MHRA) : All metal-on-metal (MoM) hip replacements: updated advice for follow-up of patients. 2017.

Milošev I, Trebše R, Kovač S, Cör A, Pišot V. Survivorship and Retrieval Analysis of Sikomet Metal-on-Metal Total Hip Replacements at a Mean of Seven Years. J Bone Jt Surg. 2006;88(6):1173–82.

Mittal S, Revell M, Barone F, Hardie DL, Matharu GS, Davenport AJ, Martin RA, Grant M, Mosselmans F, Pynsent P. Lymphoid aggregates that resemble tertiary lymphoid organs define a specific pathological subset in metal-on-metal hip replacements. PLoS One. Public Library of Science; 2013;8(5):e63470.

Moonesinghe R, Khoury MJ, Janssens ACJW. Most Published Research Findings Are False—But a Little Replication Goes a Long Way. PLoS Med. Public Library of Science; 2007;4(2):e28.

Morlock MM, Bishop N, Rüther W, Delling G, Hahn M. Biomechanical, morphological, and histological analysis of early failures in hip resurfacing arthroplasty. Proc Inst Mech Eng Part H J Eng Med. 2006;220(2):333–44.

Morlock MM, Bishop N, Zustin J, Hahn M, Rüther W, Amling M. Modes of Implant Failure After Hip Resurfacing: Morphological and Wear Analysis of 267 Retrieval Specimens. J Bone Jt Surgery-American Vol. 2008;90(Suppl 3):89–95.

Morrell AP, Floyd H, W. Mosselmans JF, Grover LM, Castillo-Michel H, Davis ET, Parker JE, Martin RA, Addison O. Improving our understanding of metal implant failures: Multiscale chemical imaging of exogenous metals in ex-vivo biological tissues. Acta Biomater. 2019;

Murgatroyd SE. Pseudotumor Presenting as a Pelvic Mass: A Complication of

Eccentric Wear of a Metal on Polyethylene Hip Arthroplasty. J Arthroplasty. Elsevier; 2012;27(5):820.e1-820.e4.

Murray DW, Grammatopoulos G, Pandit H, Gundle R, Gill HS, McLardy-Smith P. The ten-year survival of the Birmingham hip resurfacing: An independent series. J Bone Jt Surg. 2012a;94-B(9):1180–6.

Murray DW, Grammatopoulos G, Pandit H, Gundle R, Gill HS, McLardy-Smith P. The ten-year survival of the Birmingham hip resurfacing: An independent series. J Bone Jt Surg. 2012b;94(9):1180–6.

Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol A V. Adverse reactions to metal debris: histopathological features of periprosthetic soft tissue reactions seen in association with failed metal on metal hip arthroplasties. J Clin Pathol. 2012;65(5):409–18.

Nawabi DH, Nassif NA, Do HT, Stoner K, Elpers M, Su EP, Wright T, Potter HG, Padgett DE. What causes unexplained pain in patients with metal-on metal hip devices? A retrieval, histologic, and imaging analysis. Clin Orthop Relat Res. 2014;472(2):543–54.

NJR. National Joint Registry of England and Wales. 14th Annual Report 2017. 2017.

Ollivere B, Darrah C, Barker T, Nolan J, Porteous MJ. Early clinical failure of the Birmingham metal-on-metal hip resurfacing is associated with metallosis and soft-tissue necrosis. J Bone Jt Surg. 2009;91(8):1025–30.

Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gibbons CL, Ostlere S, Athanasou N, Gill HS, Murray DW. Pseudotumours associated with metal-on-metal hip resurfacings. J Bone Joint Surg Br. 2008a;90(7):847–51.

Pandit H, Vlychou M, Whitwell D, Crook D, Luqmani R, Ostlere S, Murray DW, Athanasou NA. Necrotic granulomatous pseudotumours in bilateral resurfacing hip arthoplasties: evidence for a type IV immune response. Virchows Arch. Springer-Verlag; 2008b;453(5):529.

Park S-H, Lu Z, Hastings RS, Campbell PA, Ebramzadeh E. Five Hundred Fifty-five Retrieved Metal-on-metal Hip Replacements of a Single Design Show a Wide Range of Wear, Surface Features, and Histopathologic Reactions. Clin Orthop Relat Res. 2018;476(2):261–78.

Paukkeri E-L, Korhonen R, Hämäläinen M, Pesu M, Eskelinen A, Moilanen T, Moilanen E. The Inflammatory Phenotype in Failed Metal-On-Metal Hip Arthroplasty Correlates with Blood Metal Concentrations. PLoS One. Public Library of Science; 2016;11(5):e0155121.

Paustenbach DJ, Tvermoes BE, Unice KM, Finley BL, Kerger BD. A review of the health hazards posed by cobalt. Crit. Rev. Toxicol. Taylor & Francis; 2013. p. 316–62.

Pelt CE, Erickson J, Clarke I, Donaldson T, Layfield L, Peters CL. Histologic, Serologic, and Tribologic Findings in Failed Metal-on-Metal Total Hip Arthroplasty. J Bone Jt Surg. 2013;95(21):e163.

Petit A, Mwale F, Zukor DJ, Catelas I, Antoniou J, Huk OL. Effect of cobalt and chromium ions on bcl-2, bax, caspase-3, and caspase-8 expression in human U937 macrophages. Biomaterials. Elsevier; 2004;25(11):2013–8.

Phillips EA, Klein GR, Cates HE, Kurtz SM, Steinbeck M. Histological characterization of periprosthetic tissue responses for metal-on-metal hip replacement. J Long Term Eff Med Implants. Begel House Inc.; 2014;24(1):13–23.

Pourzal R, Urban RM, Wimmer MA. What Do the Retrievals Really Tell Us. Met Bear a Clin Pract. 2013.

Reinisch G, Huber M, Zweymüller K. Verschleißtechnische Analysen und histologische Befunde bei Low-Carbon-Metall-Metall-Artikulationen des Hüftgelenks nach 11 bis 17 Jahren. Z Orthop Unfall. 2015;153(04):392–8.

Reinisch G, Judmann KP, Lhotka C, Lintner F, Zweymüller KA. Retrieval study of uncemented metal–metal hip prostheses revised for early loosening. Biomaterials. 2003;24(6):1081–91.

Reito A, Elo P, Nieminen J, Puolakka T, Eskelinen A. Gluteal muscle fatty atrophy is not associated with elevated blood metal ions or pseudotumors in patients with a unilateral metal-on-metal hip replacement. Acta Orthop. Taylor & Francis; 2016a;87(1):29–35.

Reito A, Elo P, Puolakka T, Pajamäki J, Eskelinen A. Femoral diameter and stem type are independent risk factors for ARMD in the large-headed ASR THR group. BMC Musculoskelet Disord. BioMed Central; 2015a;16(1).

Reito A, Lainiala O, Elo P, Eskelinen A. Prevalence of Failure due to Adverse Reaction to Metal Debris in Modern, Medium and Large Diameter Metal-on-Metal Hip Replacements--The Effect of Novel Screening Methods: Systematic Review and Metaregression Analysis. PLoS One. Public Library of Science; 2016b;11(3):e0147872.

Reito A, Lehtovirta L, Lainiala O, Mäkelä K, Eskelinen A. Lack of evidence—the anti-stepwise introduction of metal-on-metal hip replacements. Acta Orthop. Taylor & Francis; 2017;88(5):478–83.

Reito A, Parkkinen J, Puolakka T, Pajamäki J, Eskelinen A. Diagnostic utility of joint fluid metal ion measurement for histopathological findings in metal-on-metal hip replacements. BMC Musculoskelet Disord. BioMed Central; 2015b;16(1):1.

Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. High prevalence of adverse reactions to metal debris in small-headed ASR™ hips. Clin Orthop Relat Res. Springer; 2013;471(9):2954–61.

Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. Outcome of Birmingham hip resurfacing at ten years: role of routine whole blood metal ion measurements in screening for pseudotumours. Int Orthop. Springer; 2014;38(11):2251–7.

Ricciardi BF, Nocon AA, Jerabek SA, Wilner G, Kaplowitz E, Goldring SR, Purdue PE, Perino G. Histopathological characterization of corrosion product associated adverse local tissue reaction in hip implants: a study of 285

cases. BMC Clin Pathol. BioMed Central; 2016;16(1):1.

Rieker C, Köttig P. In Vivo Tribological Performance of 231 Metal-On-Metal Hip Articulations. HIP Int. 2002;12(2):73–6.

Salvati EA, Betts F, Doty SB. Particulate Metallic Debris in Cemented Total Hip Arthroplasty. Clin. Orthop. Relat. Res. Aug, 1993 p. 160–73.

Savarino L, Granchi D, Ciapetti G, Cenni E, Nardi Pantoli A, Rotini R, Veronesi CA, Baldini N, Giunti A. Ion release in patients with metal-on-metal hip bearings in total joint replacement: A comparison with metal-on-polyethylene bearings. J Biomed Mater Res. John Wiley & Sons, Ltd; 2002;63(5):467–74.

Scharf B, Clement CC, Zolla V, Perino G, Yan B, Elci SG, Purdue E, Goldring S, Macaluso F, Cobelli N, Vachet RW, Santambrogio L. Molecular analysis of chromium and cobalt-related toxicity. Sci Rep. Nature Publishing Group; 2014;4:5729.

Schmalzried TP, Fowble V a., Ure KJ, Amstutz HC. Metal on Metal Surface Replacement of the Hip. Clin. Orthop. Relat. Res. 1996. p. S106–14.

Scholes SC, Hunt BJ, Richardson VM, Langton DJ, Smith E, Joyce TJ. Explant analysis of the Biomet Magnum/ReCap metal-on-metal hip joint. Bone Joint Res. 2017;6(2):113–22.

Schreiber JB. Latent Class Analysis: An example for reporting results. Res Soc Adm Pharm. Elsevier; 2017;13(6):1196–201.

Seppänen M, Karvonen M, Virolainen P, Remes V, Pulkkinen P, Eskelinen A, Liukas A, Mäkelä KT. Poor 10-year survivorship of hip resurfacing arthroplasty. Acta Orthop. Taylor & Francis; 2016;87(6):554–9.

Sidaginamale RP, Joyce TJ, Bowsher JG, Lord JK, Avery PJ, Natu S, Nargol A V, Langton DJ. The clinical implications of metal debris release from the taper junctions and bearing surfaces of metal-on-metal hip arthroplasty: joint fluid and blood metal ion concentrations. Bone Joint J. 2016;98-B(7):925–33.

Sidaginamale RP, Joyce TJ, Lord JK, Jefferson R, Blain PG, Nargol AVF, Langton DJ. Blood metal ion testing is an effective screening tool to identify poorly performing metal-on-metal bearing surfaces. Bone Joint Res. 2013;2(5):84–95.

Sieber H-P, Rieker CB, Köttig P. Analysis of 118 second-generation metal-on-metal retrieved hip implants. J Bone Joint Surg Br. 1999;81-B(1):46–50.

Singh G, Meyer H, Ruetschi M, Chamaon K, Feuerstein B, Lohmann CH. Large-diameter metal-on-metal total hip arthroplasties: A page in orthopedic history? J. Biomed. Mater. Res. - Part A. 2013. p. 3320–6.

Siopack JS, Jergesen HE. Total Hip Arthroplasty. West J Med. 1995;162:243–9.

Smeekes C, Cleven AHG, van der Wal BCH, Dubois S V., Rouse RW, Ongkiehong BF, Wolterbeek R, Nelissen RGHH. Current Pathologic Scoring Systems for Metal-on-metal THA Revisions are not Reproducible. Clin Orthop Relat Res. Springer US; 2017;475(12):3005–11.

De Smet K, De Haan R, Calistri A, Campbell PA, Ebramzadeh E, Pattyn C, Gill HS. Metal Ion Measurement as a Diagnostic Tool to Identify Problems with Metal-on-Metal Hip Resurfacing. J Bone Jt Surg Am. 2008;90:202–8.

Smith & Nephew Orthopaedics. R3 Medical Device Recall. 2012.

Smith MD. The normal synovium. Open Rheumatol J. Bentham Science Publishers; 2011;5:100–6.

Steffen JE, Fassler EA, Reardon KJ, Egilman DS. Grave fraudulence in medical device research: a narrative review of the PIN seeding study for the Pinnacle hip system. Account Res. Taylor & Francis; 2018;25(1):37–66.

Van Der Straeten C, Grammatopoulos G, Gill HS, Calistri A, Campbell P, De Smet KA. The 2012 Otto Aufranc Award: The interpretation of metal ion levels in unilateral and bilateral hip resurfacing. Clin Orthop Relat Res. 2013;471(2):377–85.

Takamura KM, Amstutz HC, Lu Z, Campbell PA, Ebramzadeh E. Wear Analysis of 39 Conserve Plus Metal-on-Metal Hip Resurfacing Retrievals. J Arthroplasty. 2014;29(2):410–5.

Tarity TD, Koch CN, Burket JC, Wright TM, Westrich GH. Fretting and Corrosion at the Backside of Modular Cobalt Chromium Acetabular Inserts: A Retrieval Analysis. J Arthroplasty. 2017;32(3):1033–9.

Teo ZWW, Schalock PC. Hypersensitivity reactions to implanted metal devices: Facts and fictions. J Investig Allergol Clin Immunol. 2016;26(5):279–94.

TGA. Australian Government, Therapeutic Goods Administration (TGA). Metal-on-metal hip replacement implants: Information for general practicioners, orthopaedic surgeons and other healt professionals. 2017.

THL. Tilastoraportti: Lonkan ja polven tekonivelet 2018. Julkari. 2018.

Toms AP, Marshall TJ, Cahir J, Darrah C, Nolan J, Donell ST, Barker T, Tucker JK. MRI of early symptomatic metal-on-metal total hip arthroplasty: a retrospective review of radiological findings in 20 hips. Clin Radiol. W.B. Saunders; 2008;63(1):49–58.

Tuke MA, Scott G, Roques A, Hu XQ, Taylor A. Design Considerations and Life Prediction of Metal-on-Metal Bearings: The Effect of Clearance. J Bone Jt Surgery-American Vol. 2008;90(Suppl 3):134–41.

Uchihara Y, Grammatopoulos G, Munemoto M, Matharu G, Inagaki Y, Pandit H, Tanaka Y, Athanasou NA. Implant failure in bilateral metal-on-metal hip resurfacing arthroplasties: a clinical and pathological study. J Mater Sci Mater Med. Springer US; 2018;29(3):28.

Underwood R, Matthies A, Cann P, Skinner JA, Hart AJ. A comparison of explanted Articular Surface Replacement and Birmingham Hip Resurfacing components. J Bone Joint Surg Br. 2011;93-B(9):1169–77.

Urish KL, Giori NJ, Lemons JE, Mihalko WM, Hallab N. Trunnion Corrosion in Total Hip Arthroplasty—Basic Concepts. Orthop Clin North Am. 2019;50(3):281–8.

Vaculova J, Gallo J, Hurnik P, Motyka O, Goodman SB, Dvorackova J. Low intrapatient variability of histomorphological findings in periprosthetic tissues from revised metal/ceramic on polyethylene joint arthroplasties. J Biomed Mater Res Part B Appl Biomater. John Wiley & Sons, Ltd; 2018;106(5):2008–

18.

Vendittoli P-A, Roy A, Mottard S, Girard J, Lusignan D, Lavigne M. Metal ion release from bearing wear and corrosion with 28 mm and large-diameter metal-on-metal bearing articulations A FOLLOW-UP STUDY. J Bone Jt Surgery, Br Vol. 2010;92(1):12–9.

Whitehouse MR, Endo M, Zachara S, Nielsen TO, Greidanus N V, Masri BA, Garbuz DS, Duncan CP. Adverse local tissue reactions in metal-on-polyethylene total hip arthroplasty due to trunnion corrosion: the risk of misdiagnosis. Bone Joint J. 2015;97-B(8):1024–30.

WHO. World Health Organization. Trace elements in human nutrition and health. 1996.

Willert H-G, Buchhorn GH, Fayyazi A, Flury R, Windler M, Köster G, Lohmann CH, KÃ¶ster G, Lohmann CH. Metal-on-Metal Bearings and Hypersensitivity in Patients with Artificial Hip Joints. J Bone Jt Surg. 2005;87(1):28–36.

Willert HG, Buchhorn GH, Gobel D, Koster G, Schaffner S, Schenk R, Semlitsch M. Wear behavior and histopathology of classic cemented metal on metal hip endoprostheses. Clin Orthop Relat Res. 1996;(329 Suppl):S160-86.

Willert HG, Semlitsch M. Reactions of the articular capsule to wear products of artificial joint prostheses. J Biomed Mater Res. 1977;11:157.164.

Wimmer MA, Fischer A, Büscher R, Pourzal R, Sprecher C, Hauert R, Jacobs JJ. Wear mechanisms in metal-on-metal bearings: The importance of tribochemical reaction layers. J Orthop Res. John Wiley & Sons, Ltd; 2009;28(4):n/a-n/a.

Winter G. Tissue Reactions to Metallic Wear and Corrosion Products in Human Patients. J Biomed Mater Res. 1974;8(3):11–26.

Witt F, Bosker B., Bishop N., Ettema H., Verheyen CCP., Morlock M. The Relation Between Titanium Taper Corrosion and Cobalt-Chromium Bearing Wear in Large-Head Metal-on-Metal Total Hip Prostheses. J Bone Jt Surgery-American Vol. 2014;96(18):e157-1–9.

Witzleb W-C, Hanisch U, Kolar N, Krummenauer F, Guenther K-P, W.-C. W, U. H, N. K, F. K, K.-P. G. Neo-capsule tissue reactions in metal-on-metal hip arthroplasty. Acta Orthop. Taylor & Francis; 2007;78(2):211–20.

Witzleb W-C, Hanisch U, Ziegler J, Guenther K-P. In Vivo Wear Rate of the Birmingham Hip Resurfacing Arthroplasty. J Arthroplasty. 2009;24(6):951–6.

Wynn-Jones H, Macnair R, Wimhurst J, Chirodian N, Derbyshire B, Toms A, Cahir J. Silent soft tissue pathology is common with a modern metal-on-metal hip arthroplasty. Acta Orthop. Taylor & Francis; 2011;82(3):301–7.

Xia Z, Kwon YM, Mehmood S, Downing C, Jurkschat K, Murray DW. Characterization of metal-wear nanoparticles in pseudotumor following metal-on-metal hip resurfacing. Nanomedicine Nanotechnology, Biol Med. Elsevier Inc.; 2011;7(6):674–81.

Xia Z, Ricciardi BF, Liu Z, von Ruhland C, Ward M, Lord A, Hughes L, Goldring

SR, Purdue E, Murray D, Perino G. Nano-analyses of wear particles from metal-on-metal and non-metal-on-metal dual modular neck hip arthroplasty. Nanomedicine Nanotechnology, Biol Med. The Authors; 2017;13(3):1205–17.

# PUBLICATIONS

# PUBLICATION
# I

**Analysis of bearing wear, whole blood and synovial fluid metal ion concentrations and histopathological findings in patients with failed ASR hip resurfacings**

Lehtovirta, L., Reito, A., Parkkinen, J., Hothi, H., Henckel, J., Hart, A., & Eskelinen, A.

BMC Musculoskeletal Disorders

# Analysis of bearing wear, whole blood and synovial fluid metal ion concentrations and histopathological findings in patients with failed ASR hip resurfacings

Lari Lehtovirta[1,2*] , Aleksi Reito[2], Jyrki Parkkinen[3], Harry Hothi[4], Johann Henckel[4], Alister Hart[4] and Antti Eskelinen[2]

## Abstract

**Background:** Adverse Reaction to Metal Debris (ARMD) is still a major reason for revision surgeries in patients with metal-on-metal (MoM) hip replacements. ARMD consists of a wide range of alterations in periprosthetic tissues, most important of which are metallosis, inflammation, pseudotumors and necrosis. Studies investigating histopathological findings and their association to implant wear or indirect measures of wear have yielded inconsistent results. Therefore, we aimed to investigate bearing surface wear volume, whole blood and synovial fluid metal ion concentrations, histopathological findings in periprosthetic tissues and their associations.

**Methods:** Seventy-eight patients with 85 hips revised for ARMD were included in the study. Prior to revision surgery, all patients had whole blood chromium and cobalt ion levels assessed. In revision surgery, a synovial fluid sample was taken and analyzed for chromium and cobalt. Periprosthetic tissue samples were taken and analyzed for histopathological findings. Explanted implants were analyzed for bearing wear volume of both acetabular cup and femoral head components.

**Results:** Volumetric wear of the failed components was highly variable. The total wear volume of the head and cup had a strong correlation with whole blood chromium and cobalt ion concentrations (Cr: $\rho = 0.80$, $p < 0.001$ and Co: $\rho = 0.84$, $p < 0.001$) and a bit weaker correlation with fluid chromium and cobalt ion concentrations (Cr: $\rho = 0.50$, $p < 0.01$ and Co: $\rho = 0.41$, $p = 0.027$). Most tissues displayed only low-to-moderate amounts of macrophages and lymphocytes. Total wear volume correlated with macrophage sheet thickness ($\rho = 0.25$, $p = 0.020$) and necrosis ($\rho = 0.35$, $p < 0.01$). Whole blood chromium and cobalt ion concentrations had similar correlations. Lymphocyte cuff thickness did not correlate with either total wear volume or whole blood metal ion concentrations, but correlated with the grade of necrosis.

**Conclusions:** Bearing wear volume correlated with blood metal ion levels and the degree of necrosis and macrophage infiltration in periprosthetic tissues suggesting a dose-response relationship. Whole blood metal ion levels are a useful tool for clinician to estimate bearing wear and subsequent tissue response.

**Keywords:** Metal-on-metal hip replacement, Adverse reaction to metal debris, ARMD, ALTR, ALVAL, Wear, Histopathology

* Correspondence: lehtovirta.lari.a@student.uta.fi
[1]Faculty of Medicine, University of Tampere, Tampere, Finland
[2]Coxa Hospital for Joint Replacement, Tampere, Finland
Full list of author information is available at the end of the article

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 2 of 11

# Background

Adverse reaction to metal debris (ARMD) is still a major reason for revision surgeries in patients with metal-on-metal (MoM) hip implants. Although the use of MoM hip implants has been widely ceased, more than one million patients have received such a device [1] and those that have not been revised still pose an increased risk for implant failure. ARMD is an umbrella term describing a wide range of alterations seen macro- and microscopically in the periprosthetic tissue such as metallosis, necrosis, inflammation of different types and soft-tissue inflammatory lesions referred as pseudotumors [2–4].

Retrieval studies have investigated implant wear and its association to ARMD. Results of these studies have been inconclusive as adverse reactions have been observed both in patients with high and low wearing hip implants [5–11]. In their recent systematic review, Campbell et al. concluded that no clear dose-response relationship between wear and ARMD could be established due to the heterogeneity of the findings in the included studies. Studies that have investigated wear or indirect markers of wear, such as synovial fluid (SF) or whole blood (WB) metal ion concentrations, and the histopathological features of ARMD have also yielded inconsistent results [8, 9, 12–19]. Extra-articular tissues retrieved from patients with ARMD vary considerably in their histologic presentation. Most often tissues display prominent macrophage infiltration as a response to the cytotoxic metal wear debris with a variable amount of lymphocytic infiltration, either diffuse or aggregated [2, 3, 8, 13]. However, in a minority of patients with ARMD, there is heavy lymphocytic infiltration, resembling type IV hypersensitivity reaction [17, 20–23]. The presence of lymphocytes is usually accompanied with the presence of necrosis and this type of tissue response was first termed ALVAL (Aseptic Lymphocytic Vasculitis-Associated Lesion) by Willert et al. [20]. Terms ALVAL and ARMD have however been inappropriately used as synonyms in the recent literature [24]. Low bearing wear has been associated with a suspected metal hypersensitivity response in some studies [8, 9, 16]. Vice versa, high bearing wear has been associated with a macrophage-dominated foreign-body response. [8, 13]. In addition, low WB metal ion levels have been associated with lymphocyte-dominated tissue response and high metal ion levels with macrophage-dominated response [17]. Based on these findings, metal hypersensitivity to implant-derived debris has been hypothesized as a cause of ARMD in patients with low-wearing hip implants, and cytotoxic, macrophage dominated response in patients with high-wearing hip implants [8, 13, 16, 17] However, findings not supporting these hypotheses have been published as well [10, 12, 15, 18, 19].

The histopathology of ARMD has been well described but the literature regarding its association to implant wear is inconsistent. It is important to understand the true nature of the association between wear and histopathological findings in ARMD. After all, it is the histopathological changes – tissue destruction and inflammation – that lead to failure of MoM hip implants. Implant wear cannot be measured in-vivo and thus cannot be used in clinical decision making but there are reliable indirect measures of wear, such as WB metal ion levels, that are commonly used in the follow-up of patients with MoM hip replacements. To gain a better understanding of the relationships between histopathological findings, bearing wear and clinical markers of wear we aimed to investigate bearing surface wear volume, WB and SF metal ion concentrations as clinical markers of wear, and their associations with histopathological findings of the periprosthetic tissue in patients with Articular Surface Replacement (ASR) hip resurfacing device revised due to ARMD. Based on the previous literature we hypothesized that 1) low implant wear is associated with high amount of lymphocytes characteristic of an ALVAL response and 2) high implant wear is associated with high amount of macrophages characteristic of a foreign-body response to metal wear debris.

# Methods

Between the recall of the ASR MoM hip system (Depuy Orthopaedics, Warsaw, IN, USA) in August 2010 and the end of our recruitment period in January 2016, 114 ASR hip resurfacing devices in 107 patients have been revised at our institution. All consecutively revised patients who gave informed consent and fulfilled the following criteria were included in our study: 1) Revision was due to ARMD, 2) Retrieved components were available for bearing wear analysis and 3) Periprosthetic tissue sample was available for histopathologic analysis. After exclusion, 85 hips in 78 patients were included in our study. Twenty-one of these patients were referred to our institution from central hospitals from other hospital districts and 57 patients had had their index operation (primary arthroplasty) and follow-up at our institution. Surgery was performed by or under the direct supervision of 14 senior orthopaedic surgeons. The study was approved by the ethical committee of Pirkanmaa Hospital District (R11006).

Revision surgery was considered if 1) a clear pseudotumour (Imperial class 2A,2B or 3) [25] was observed on cross-sectional imaging regardless of symptoms or WB metal ion levels; or 2) the patient had elevated WB metal ion levels and hip symptoms despite normal findings in cross-sectional imaging; or 3) the patient had a continuously symptomatic hip or progressive symptoms regardless

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 3 of 11

of imaging findings or metal ion levels. Symptoms included hip pain, discomfort, sense of instability, and/or impaired function of the hip and sounds from the hip (clacking, squeaking). WB metal ion levels were regarded as being elevated if either chromium or cobalt exceeded 5 ppb. Post-operatively, failure was classified as being due to ARMD on the basis of the following criteria [26, 27]: 1) there was presence of metallosis or macroscopic synovitis in the joint; and/or 2) a pseudotumor was found during revision; and/or 3) a moderate to high number of perivascular lymphocytes along with tissue necrosis and/or fibrin deposition was seen in the histopathologic sample; and 4) perioperatively there was no evidence of component loosening or periprosthetic fracture. In addition, infection was ruled out by obtaining multiple (at least five) bacterial cultures during revision surgery.

### Bearing wear analysis

The volume of material loss from the cup and head bearing surfaces was measured using a Zeiss Prismo (Carl Zeiss Ltd., Rugby, UK) coordinate measuring machine (CMM). A total of 400 polar scan lines on each surface were defined and up to 30,000 data points captured using a 2 mm ruby stylus; protocols for this method have been previously published [28]. An iterative least square fitting method was used to analyse the raw data captured by the CMM and the unworn geometry of the bearing surface was used to map regions of material loss from which the total volumetric loss was calculated. Wear rate (mm3/year) was further calculated by dividing total wear volume in cubic millimeters by implantation time in years.

### Histopathological analysis of the periprosthetic tissue

During every hip revision, a sample of the inflamed synovia or pseudotumor was obtained. For histopathological analysis, each tissue sample was formalin fixed. Several 10 μm microtome sections were made and embedded in paraffin. Standard hematoxylin and eosin staining was used. The sections were examined histologically under normal light with a Nikon Eclipse 50i (Nikon Corporation, Shinagawa, Tokyo, Japan). The samples were graded by a senior musculoskeletal pathologist (JP) using scoring principles adopted from the study by Natu et al. [2] (termed Natu grading in our study) and the ALVAL score previously described by Campbell et al. [8].

The Natu grading consisted of following parameters: 1) lymphocyte cuff thickness, 2) whether lymphocytic infiltrate was diffuse or aggregated, 3) presence of germinal centers, 4) histiocyte sheet thickness, 5) metal particle load within histiocytes, 6) extent of tissue necrosis, 7) presence of plasma cells and 8) presence of granulomas. Lymphocyte cuff thickness was calculated using a

graticule. An average of five measurements was taken and graded as 0–3 (absent, 0.25 mm, 0.25–0.75 mm, >0.75 mm). Thickness of histiocyte sheets was also calculated using a graticule and graded 0–3 (absent, <1 mm, 1–2 mm, >2 mm). Metal particle load within histiocytes was graded as 0–4 as done in the assessment of iron decomposition in liver cells [29, 30]. The extent of overall tissue necrosis in a sample was graded based on the surface necrosis typing according to Davies et al. [22]. Type 1 surface contains intact synovial epithelium. Type 2 surface shows loss of synovial epithelial cells without fibrin deposition. In type 3 surface there is fibrin deposition and in type 4 surface there is extensive necrosis and loss of architecture. The extent of type 4 surface necrosis was used to grade the overall tissue necrosis in a given sample, as described by Natu et al. [2]. In grade 4 necrosis, more than 75% of the tissue sample showed type 4 surface necrosis. In grade 3 necrosis, between 25 and 75% showed type 4 surface necrosis. In grade 2 necrosis either less than 25% of the tissue showed type 4 surface necrosis or the tissue showed type 3 surface. In grade 1 necrosis, the sample consisted of type 2 surface.

ALVAL scoring consists of three subscores: synovial lining (0-3p), tissue organization (0-3p) and inflammatory infiltrate (0-4p). Both synovial lining and tissue organization reflect the degree of necrosis and higher scores mean higher degree of necrosis. Inflammatory infiltrate score reflects the predominant inflammatory cell type on a spectrum: 0 points means minimal infiltrates, 1p means predominantly macrophages, 2p means both macrophages and diffuse/perivascular lymphocytes, 3p means mostly lymphocytes in aggregates and some macrophages and 4p means large lymphocyte aggregates and little to no macrophages.

### Whole blood and synovial fluid metal analysis

Since January 2012, WB metal ion (Co and Cr) concentrations have been routinely measured as a part of the systematic follow-up program for patients with MoM hip replacements at our institution. All patients underwent WB analysis of Co/Cr following sampling from the antecubital vein using a 21-gauge needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) and trace-element blood tubes containing sodium ethylenediaminetetraacetic acid (EDTA). Standard operating procedures were established at the Finnish Institute for Occupational Health for Co and Cr measurement using dynamic reaction cell inductively coupled plasma (quadripole) mass spectrometry (Agilent 7500 cx, Agilent Technologies, Santa Clara, CA, USA). The laboratory technicians were blinded to all clinical outcomes. The samples were preserved in +6 °C to +8 °C prior to analysis.

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 4 of 11

Since October 2011, our MoM hip revision protocol has involved perioperative SF aspiration, which is always taken before opening the deep fascia using a standard 18- to 20-gauge needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, New Jersey) and trace element tubes containing sodium EDTA. Similar procedures were used for SF metal ion concentration measurement as described above for WB.

### Statistical methods

Spearman rank correlation was used to study the associations between wear volume, WB and SF metal ion concentrations, and histopathological findings due to non-normal distribution of these variables. Medians were calculated for wear volume, WB and SF metal ion concentrations. To compare these median values between different subgroups, nonparametric Mann-Whitney U-test was used. When analyzing the correlation between WB metal ion concentrations and other factors, we only included patients with unilateral hip arthroplasties (57 hips) to avoid the confounding effect of metal ions being released to the blood from the other implant. The threshold for statistical significance was set to 0.05. The analyses were conducted using IBM SPSS software (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.).

### Results

Of the 85 hips included in the study, 56 were explanted from female patients and 29 from male patients. Mean age at the time of the revision surgery was 57.3 years (SD 10.3 years). Mean follow-up time between index operation and revision surgery was 5.4 years (SD 1.8 years).

Volumetric wear analysis of the explanted components demonstrated a wide range of wear in both the acetabular cup and femoral head (Table 1). Wear rates were also highly variable with a median of 9.0 mm3/year (range 1.1...99.7 mm3/year). In a vast majority of the components (85.1%), the femoral head was more worn than the acetabular cup. Median ratio for head wear to cup wear was 1.7 (range, 0.5...10). In addition to actual volumetric component wear, also WB and SF metal ion levels, serving as indirect markers of wear, were highly variable (Table 2). The total wear volume of the head and cup strongly correlated with WB metal ion concentrations (Cr: $\rho = 0.80$, $p < 0.001$ and Co: 0.84, $p < 0.001$) and

**Table 1** Median volumetric wear and range for acetabular and femoral components and both combined

| Component | Median volumetric wear (mm$^3$) | Range (mm$^3$) |
|---|---|---|
| Acetabular cup | 14 | 2–247 |
| Femoral head | 24 | 4–485 |
| Both combined | 39 | 7–541 |

**Table 2** Median concentrations (μg/l) and ranges (μg/l) for chromium and cobalt ions in both whole blood and synovial fluid

| Metal ion | Whole blood | Range | Synovial fluid | Range |
|---|---|---|---|---|
| Chromium | 9.7 | 0.5–93.9 | 701 | 7.0–52360 |
| Cobalt | 15.4 | 0.7–224.7 | 281.5 | 27.0–14870 |

moderately with SF metal ion concentrations (Cr: $\rho = 0.50$, $p < 0.01$ and Co: $\rho = 0.41$, $p = 0.027$). Wear rate had slightly stronger correlation with WB metal ion concentrations (Cr: $\rho = 0.87$, $p < 0.001$ and Co: 0.89, $p < 0.001$) and SF metal ion concentrations (Cr: $\rho = 0.71$, $p < 0.001$ and Co: 0.66, $p < 0.01$) than total wear volume.

Histologically, variable amounts of macrophages, lymphocytes and necrosis were seen in the tissue samples. One or more germinal centers were present in 5 samples (5.7% of all samples). One or more granulomas were present in 14 samples (16.5% of all samples). All tissue samples evinced at least some degree of macrophage infiltration (macrophage sheet thickness score of at least 1) and in most cases it was low to moderate (Fig. 1). In regard to lymphocyte infiltration, most tissues evinced little to no lymphocytes and in only a minority of the samples the infiltrate was prominent (Fig. 2). All cases with heavy lymphocyte infiltration (scores 2 or 3) had a macrophage sheet thickness score of 1, ie. there was only little macrophage infiltration in these tissues. Eight of the nine tissue samples with heavy lymphocyte infiltration had grade 4 necrosis and the ninth had grade 3 necrosis. Median wear rate for these tissues was higher than that for the tissues with lower numbers of lymphocytes (Table 3). Lymphocyte cuff thickness correlated positively with the grade of necrosis ($\rho = 0.41$, $p < 0.001$) and inflammatory infiltrate score ($\rho = 0.79$, $p < 0.001$).. All five tissue samples with germinal centers had grade 4 necrosis. The wear volume or wear rate of these cases did not differ from cases without germinal centers (median wear volume in cubic millimeters 61 versus 37.5, $p = 0.94$; median wear rate in cubic millimeters/year 11.9 versus 8.8, $p = 0.86$). Tissues with one or more granulomas were associated with higher total wear volume and wear rate when compared to tissues with no granulomas (Table 4). Grade of necrosis correlated positively with synovial lining score ($\rho = 0.86$, $p < 0.001$) and tissue organization score ($\rho = 0.80$, $p < 0.001$).

Correlations between histological variables, total wear volume of the head and cup components, wear rate as well as WB and SF metal ion concentrations are listed in Table 5. Total wear volume correlated with macrophage sheet thickness, grade of necrosis (Fig. 3), synovial lining score, tissue organization score and total ALVAL score. Wear rate had similar correlations but the correlation with macrophages did not quite reach statistical significance. WB cobalt and chromium ion concentrations had

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523
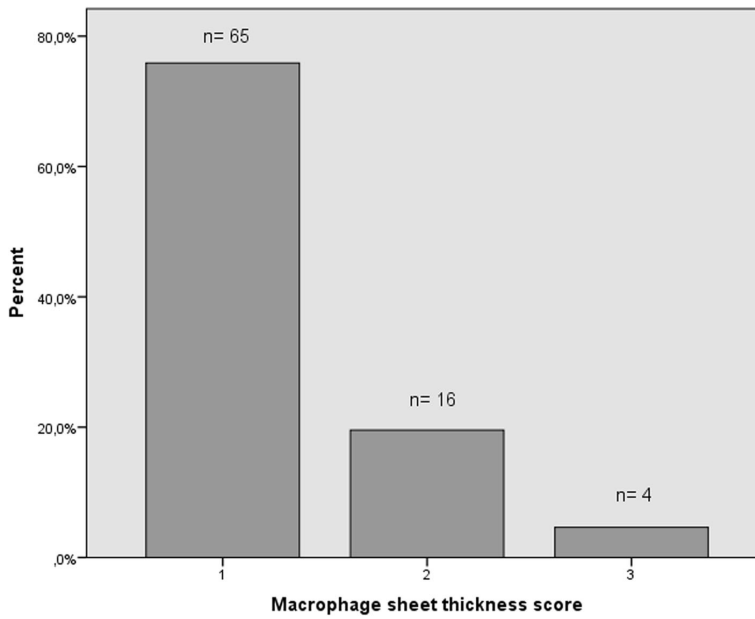
Page 5 of 11



**Fig. 1** Distribution of macrophage sheet thickness scores among all periprosthetic tissue samples
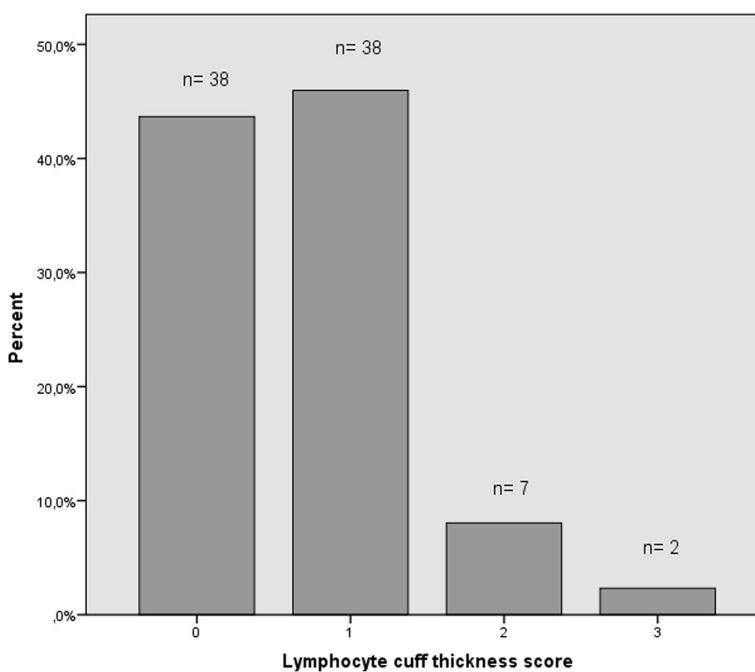


**Fig. 2** Distribution of lymphocyte cuff thickness scores among all periprosthetic tissue samples

Lehtovirta *et al. BMC Musculoskeletal Disorders*  (2017) 18:523

Page 6 of 11

**Table 3** Wear rates according to lymphocyte cuff thickness

|  | Lymphocyte cuff <2 | Lymphocyte cuff 2 or 3 | *P*-value |
|---|---|---|---|
| Median wear rate (mm3/year) | 8.1 | 14.5 | 0.054 |
| Range (mm3/year) | 1.1 … 99.8 | 3.7 … 48.0 | |

similar correlations. SF chromium ion concentration correlated with grade of necrosis, synovial lining score and total ALVAL score. SF cobalt ion concentration correlated with all but macrophage sheet thickness. Neither wear volume, wear rate, WB metal ion concentrations or SF chromium ion concentration were associated with lymphocyte cuff thickness or presence of germinal centers. However, SF cobalt ion concentration did correlate with lymphocyte cuff thickness.

## Discussion

In the present study, a spectrum of inflammatory and necrotic changes associated with ARMD were seen – variable macrophage and lymphocyte infiltration and necrosis in periprosthetic tissues [2–4, 8, 21]. Most patients evinced low-to-moderate macrophage infiltration and little to no lymphocyte infiltration. A few patients evinced a very prominent lymphocyte infiltration with grade 4 necrosis typical of an ALVAL response first proposed by Willert et al. [20]. The thickness of lymphocyte cuffs correlated positively with the degree of necrosis. Bearing wear and WB metal ion concentrations correlated positively with the number of macrophages and the degree of necrosis.

Our study is not without limitations. Firstly, due to the periprosthetic tissue being sampled only at the time of revision surgery, it is difficult to say about the natural history of ARMD. Secondly, tissue samples were analyzed by one observer only. However, multiple microtome sections of each sample were made and analyzed by a senior musculoskeletal pathologist well acquainted with ARMD histopathology. Thirdly, we did not perform a priori sample size calculation. Our study was retrospective of nature and patients were included on an "all-comer" basis. However, a posteriori power analysis

**Table 4** Wear volume and rate: comparison between patients with one or more granulomas and those without granulomas

|  | Granuloma present | Granuloma absent | *P*-value |
|---|---|---|---|
| Median total wear volume (mm3) | 106.5 | 31.0 | 0.016 |
| Range (mm3) | 10…378 | 7…541 | |
| Median wear rate (mm3/year) | 16.3 | 8.1 | 0.035 |
| Range (mm3/year) | 1.6…99.8 | 1.1…86.4 | |

revealed that our study has 90% power (10% beta) to detect 0.35 correlation (medium effect size) with a type I error probability of 5% (alfa). Fourthly, although we consecutively recruited patients, not all patients who underwent surgery because of ARMD during the recruitment period were included due to refused consent, missing tissue samples and missing wear data on some patients. Thus, our patient series is not completely consecutive. However, the number of excluded patients was low in comparison to the number of consecutive patients included. Indeed, the large number of patients included is a major strength of our study. Another strength is that we only included patients revised for ARMD and with identical hip resurfacing implants. Thus, our data specifically describes patients with ARMD while minimizing the confounding effect from having different implant designs or failure modes other than ARMD. Also, there was no confounding effect from possible trunnion wear debris as in the case of THRs since we only investigated the effects of bearing wear debris.

Metal debris originating from the bearing surfaces and/or trunnion has been shown to have cytotoxic effects [31–34]. It has been suggested that the cytotoxicity of metal debris further leads to tissue destruction and macrophage recruitment to clear the tissue and metal debris [3, 21]. In support of this, we observed a correlation between implant wear and the number of macrophages as well as the degree of necrosis but not with the number of lymphocytes. Similar findings were made in a study by Grammatopoulos et al. [13]. Langton et al. however did not find correlation between wear and the amount of macrophages or necrosis [6]. We also observed that the presence of granulomas was associated with increased total wear volume. Granulomas are thought to form in response to high number of wear particles and our results support this idea [35]. High wear, or high WB metal ion concentrations, have been associated to macrophage-dominated tissue responses in other studies as well [8, 16, 17]. Metal hypersensitivity leading to type IV response with strong lymphocytic infiltration has been suggested as a cause of failure in those patients with low wear [8, 16, 17]. Contrary to our hypothesis, this was not observed in our study. In fact, patients with heavy lymphocytic infiltration had higher wear rate than patients with lower numbers of lymphocytes. These findings suggest that excessive metal debris accumulation was the cause of lymphocytic infiltration in these patients, not metal hypersensitivity. Grammatopoulos et al. also did not find correlation between wear and lymphocytic infiltration, but noted the presence of a patient subgroup with hypersensitivity-related histopathological findings and simultaneous low bearing wear, suggesting metal hypersensitivity as a cause of failure in those patients [13].
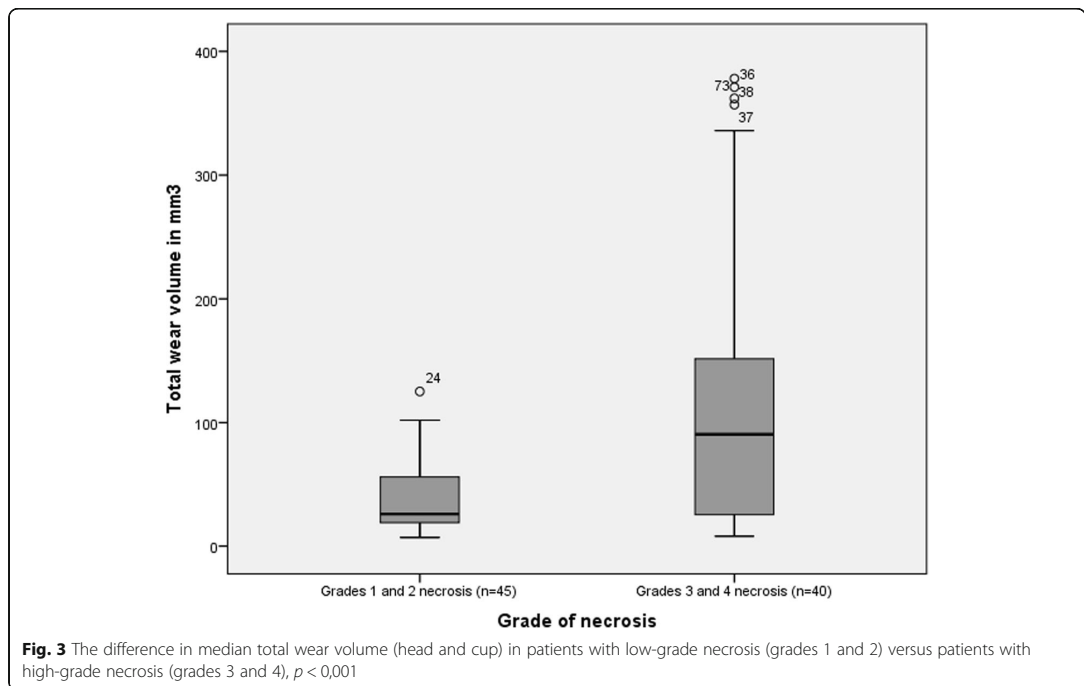
Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 7 of 11

**Table 5** Spearman rho correlation coefficients and associated *p*-values for correlations between total wear volume, wear rate, indirect markers of wear (whole blood and synovial fluid metal ion concentrations) and histopathological grading (Natu and ALVAL)

|  | Natu grading | | | ALVAL grading | | | |
|---|---|---|---|---|---|---|---|
|  | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Inflammatory infiltrate score | Synovial lining score | Tissue organization score | Total ALVAL score |
| Total wear volume | rho = 0.11<br>*p* = 0.32 | rho = 0.25*<br>*p* = 0.020 | rho = 0.35*<br>*p* < 0.01 | rho = 0.13<br>*p* = 0.23 | rho = 0.37*<br>*p* < 0.01 | rho = 0.25*<br>*p* = 0.023 | rho = 0.31*<br>*p* < 0.01 |
| Wear rate | rho = 0.17<br>*p* = 0.12 | rho = 0.20<br>*p* = 0.069 | rho = 0.42*<br>*p* < 0.0001 | rho = 0.16<br>*p* = 0.15 | rho = 0.48*<br>*p* < 0.0001 | rho = 0.35*<br>*p* < 0.01 | rho = 0.40*<br>*p* < 0.001 |
| WB Cr | rho = 0.089<br>*p* = 0.51 | rho = 0.30*<br>*p* = 0.024 | rho = 0.45*<br>*p* < 0.001 | rho = 0.19<br>*p* = 0.26 | rho = 0.54*<br>*p* < 0.001 | rho = 0.33*<br>*p* = 0.015 | rho = 0.48*<br>*p* < 0.001 |
| WB Co | rho = 0.18<br>*p* = 0.18 | rho = 0.29*<br>*p* = 0.029 | rho = 0.51*<br>*p* < 0.001 | rho = 0.26<br>*p* = 0.055 | rho = 0.60*<br>*p* < 0.001 | rho = 0.40*<br>*p* < 0.01 | rho = 0.55*<br>*p* < 0.001 |
| SF Cr | rho = 0.30<br>*p* = 0.12 | rho = 0.16<br>*p* = 0.40 | rho = 0.48*<br>*p* < 0.01 | rho = 0.25<br>*p* = 0.19 | rho = 0.56*<br>*p* < 0.01 | rho = 0.34<br>*p* = 0.070 | rho = 0.53*<br>*p* < 0.01 |
| SF Co | rho = 0.49*<br>*p* < 0.01 | rho = 0.17<br>*p* = 0.37 | rho = 0.54*<br>*p* < 0.01 | rho = 0.44*<br>*p* = 0.017 | rho = 0.47*<br>*p* = 0.011 | rho = 0.38*<br>*p* = 0.045 | rho = 0.57*<br>*p* < 0.01 |

Values that are statistically significant are flagged with *

There are several reasons that likely contribute to the inconsistency of findings between different histopathological studies. It is possible and probable that some patients have both high-wearing implants and an underlying hypersensitivity response that would have evoked even in the presence of a low-wearing implant. This combination may result in a mixed-type tissue response that has the characteristics of both wear-related foreign-body response and hypersensitivity-related type IV tissue responses and therefore makes it difficult to distinguish between the two based on implant wear or WB metal ion levels alone. Also, threshold for the onset of adaptive immune response is likely variable between individuals [13]. This would explain why some patients tolerate extensive amount of wear debris and some patients develop ALVAL in the presence of a low wearing MoM hip replacement. In addition to patient susceptibility, differences in implant types among studies may



**Fig. 3** The difference in median total wear volume (head and cup) in patients with low-grade necrosis (grades 1 and 2) versus patients with high-grade necrosis (grades 3 and 4), *p* < 0,001

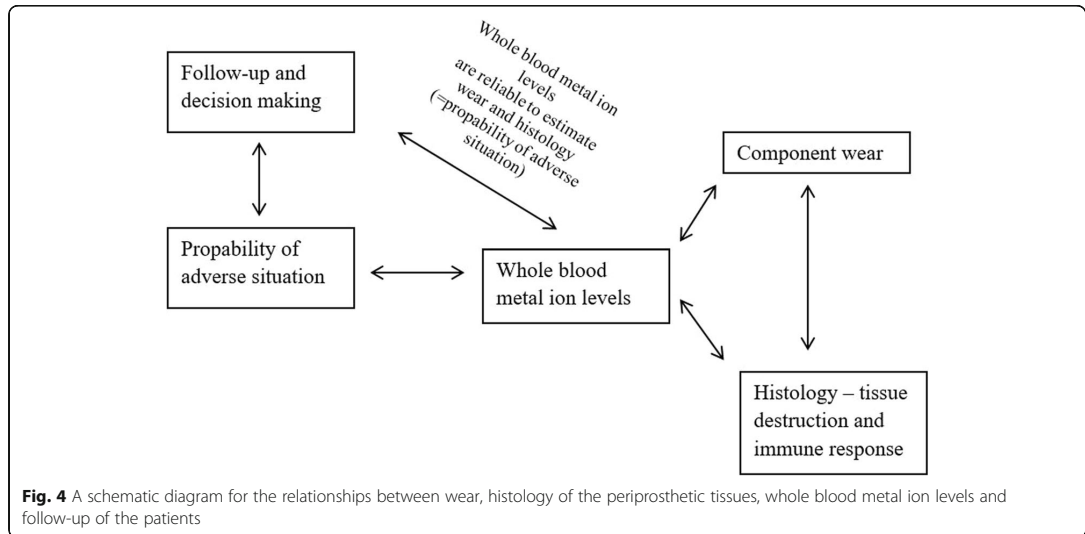Lehtovirta et al. BMC Musculoskeletal Disorders (2017) 18:523

Page 8 of 11

play a role. Substantially higher failure rates have been reported for ASR XL THRs in comparison with ASR hip resurfacings [5, 19]. ASR XL and ASR have similar bearing couples, but in the ASR XL there is trunnion-interface between the titanium stem and CoCr head that serves as an additional source of metal debris. It has been shown that wear from the trunnion is different in nature compared to bearing surface wear and may lead to lymphocytic and necrotic tissue responses [14, 36, 37], possibly contributing to the inconsistencies between other recent histopathological studies.

Natu et al. suggested that the lymphocytic immune response in patients with MoM implants is a dynamic process, beginning with perivascular lymphocytic aggregates and leading to formation of lymphoid follicles with germinal centers, also termed as tertiary lymphoid organs (TLOs) [2]. TLOs are capable of forming new B and T cells locally. TLOs are seen in affected tissues of patients with chronic autoinflammatory diseases such as rheumatoid arthritis, Sjogrens syndrome and Hashimoto's thyroiditis and are considered to be formed as a response to a persistent antigen that cannot be eliminated [38]. Mittal et al. demonstrated the presence of TLOs and associated chemokines in tissues of patients with failed MoM hip implants and suggested that these patients form a specific pathological subset [39] in addition to the well-established foreign-body response [3, 13] and ALVAL response [8, 20, 40]. In keeping with findings by Natu et al. and Mittal et al., we found that a minority of the patients displayed lymphoid follicles with germinal centers. Further, all tissue samples displaying germinal centers had grade 4 necrosis. This suggests that the formation of TLOs in periprosthetic tissues is associated with tissue destruction, possibly accelerating the process of implant failure. In line with this is a finding that the presence of TLOs has been associated with tissue destruction and loss of function in autoimmune diseases [38]. In the present study, we also observed that even in the absence of germinal center containing lymphoid follicles, the thickness of lymphocytic cuffs correlated with the grade of necrosis. Whether the lymphocytic immune response, also termed ALVAL, is a dynamic process leading to formation of TLOs as Natu et al. suggested [2] or whether patients with TLOs define their own distinct pathological subset as Mittal et al. suggested [39] requires further research. However, due to the cross-sectional nature of histopathological studies, it is difficult to investigate the natural history of ARMD.

ALVAL grading introduced by Campbell et al. [8] has been used in several studies [9, 16, 18, 41] but other grading systems have been used as well [2, 13, 21, 22, 36]. This makes comparison between studies difficult. In the present study, all tissue samples were analyzed according to two grading criteria: ALVAL grading and grading principles established by Natu et al. [2]. ALVAL grading is relatively restricted compared to the Natu grading as it only includes inflammatory infiltrate score, synovial lining score and tissue organization score. Moreover, as discussed by Ricciardi et al., both synovial lining score and tissue organization score reflect the degree of necrosis [21]. A strong correlation between these scores and the Natu score for necrosis was observed in our study. ALVAL score was originally designed to help distinguish failures related to high wear from failures related to suspected hypersensitivity (ALVAL) response. Although necrosis is often seen with ALVAL response, it is not specific for ALVAL as it is also seen with macrophage-dominated foreign body reactions with possible related cytotoxicity [3, 13]. This leaves only the inflammatory infiltrate score in ALVAL grading specific for ALVAL response. In the present study, a strong correlation between lymphocyte cuff thickness and inflammatory infiltrate score was observed. This indicates that the inflammatory infiltrate score is useful in distinguishing lymphocyte-dominated responses from those that are not lymphocyte-dominated. Phillips et al. also concluded that ALVAL scoring is useful for distinguishing between macrophage and lymphocyte responses [42]. Inflammatory infiltrate score involves evaluation of both lymphocytic and macrophagic components. However, both lymphocytes and macrophages are often seen in periprosthetic tissues. Grammatopoulos et al. suggested that an easier method to identify ALVAL responses from wear-related responses would be to measure only the thickness of lymphocytic cuffing, termed Oxford-ALVAL score in their study [13]. We agree with Grammatopoulos et al. and find that separate scores for evaluation of macrophage and lymphocyte infiltration provide more information about the failure mechanism and lead to easier comparison between histopathological studies.

In the present study, WB metal ion levels had a strong correlation with bearing wear volume, wear rate and a moderate correlation with several histopathological features. SF metal ion levels also correlated with bearing wear volume, wear rate and some of the histopathological features, but the correlations were weaker. Wear rate had a stronger correlation with WB and SF metal ion levels than total bearing volume. Wear rate, WB and SF metal ion levels likely reflect the recent burden of metal debris whereas total wear volume reflects the amount of total wear accumulated during implantation time. In a study by De Smet et al. both WB and SF metal ion levels were found to correlate well with linear wear of the femoral component [43]. Langton et al. also noted a correlation between wear volume and WB metal ion levels [6]. Interestingly, WB metal ion levels had stronger correlation with histopathological findings compared to SF metal ion levels in the present study. This is

Lehtovirta et al. BMC Musculoskeletal Disorders  (2017) 18:523

Page 9 of 11



**Fig. 4** A schematic diagram for the relationships between wear, histology of the periprosthetic tissues, whole blood metal ion levels and follow-up of the patients

supported by a recent study by Reito et al. who found that SF metal ion levels had relatively poor correlation with histopathological findings [12]. SF aspiration is an invasive procedure and the measurement does not seem to provide any additional information compared to WB measurement. Measurement of WB metal ion levels is a reliable, indirect way to gain information of the in-situ wear process, inflammation and tissue destruction. Clinicians should use WB metal ion levels in the follow-up of patients with MoM hip implants and closely monitor those patients with elevated metal ion levels. A schematic diagram for the relationships between wear, histology of the periprosthetic tissues, WB metal ion levels and follow-up of the patients is presented in Fig. 4.

## Conclusions

In the present study with failed ASR hip resurfacings, total wear volume, wear rate and WB metal ion concentrations correlated with the number of macrophages and the degree of necrosis, but not with the amount of lymphocytes. Most tissue samples evinced macrophages but little to no lymphocytes typical of a non-specific foreign-body response. A minority of the samples evinced strong lymphocyte infiltration combined with high amount of necrosis, typical of an ALVAL response. However, contrary to our hypothesis, this type of response was not associated with low implant wear in the present study. The significance of patient susceptibility in the development of ARMD is poorly understood and it is not currently known which factors lead to the adaptive lymphocytic response seen in some patients. Future studies should be directed to understand the patho-physiological mechanisms behind different types of

tissue responses seen in patients with MoM hips. WB metal ion levels correlated with total wear volume, wear rate and histopathological findings. Measurement of WB metal ion levels is useful in the follow-up of patients with hip resurfacings as it provides information of the wear process, inflammatory response and tissue destruction.

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 10 of 11

## Author details
[1]Faculty of Medicine, University of Tampere, Tampere, Finland. [2]Coxa Hospital for Joint Replacement, Tampere, Finland. [3]Fimlab Laboratories Oy, Tampere, Finland. [4]University College London, London, UK.

## References
1. Lombardi AV Jr, Barrack RL, Berend KR, Cuckler JM, Jacobs JJ, Mont MA, Schmalzried TP. The hip society: algorithmic approach to diagnosis and management of metal-on-metal arthroplasty. J Bone Joint Surg Br. 2012;94(11 Suppl A):14–8.
2. Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol AV. Adverse reactions to metal debris: histopathological features of periprosthetic soft tissue reactions seen in association with failed metal on metal hip arthroplasties. J Clin Pathol. 2012;65(5):409–18.
3. Mahendra G, Pandit H, Kliskey K, Murray D, Gill HS, Athanasou N. Necrotic and inflammatory changes in metal-on-metal resurfacing hip arthroplasties: relation to implant failure and pseudotumor formation. Acta Orthop. 2009; 80(6):653–9.
4. Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gibbons CL, Ostlere S, Athanasou N, Gill HS, Murray DW. Pseudotumours associated with metal-on-metal hip resurfacings. J Bone Joint Surg Br. 2008;90(7):847–51.
5. Langton D, Jameson S, Joyce T, Hallab N, Natu S, Nargol A. Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: a consequence of excess wear. J Bone Joint Surg Br. 2010;92(1):38–46.
6. Langton D, Joyce T, Jameson S, Lord J, Van Orsouw M, Holland J, Nargol A, De Smet K. Adverse reaction to metal debris following hip resurfacing the influence of component type, orientation and volumetric wear. J Bone Joint Surg Br. 2011;93(2):164–71.
7. Kwon Y, Glyn-Jones S, Simpson D, Kamali A, McLardy-Smith P, Gill H, Murray D. Analysis of wear of retrieved metal-on-metal hip resurfacing implants revised due to pseudotumours. J Bone Joint Surg Br. 2010;92(3):356–61.
8. Campbell P, Ebramzadeh E, Nelson S, Takamura K, De Smet K, Amstutz H. Histological features of Pseudotumor-like tissues from metal-on-metal hips. Clin Orthop Relat Res. 2010;468(9):2321–7.
9. Ebramzadeh E, Campbell PA, Takamura KM, Lu Z, Sangiorgio SN, Kalma JJ, De Smet KA, Amstutz HC. Failure modes of 433 metal-on-metal hip implants: how, why, and wear. Orthop Clin North Am. 2011;42(2):241–50.
10. Ebramzadeh E, Campbell P, Tan TL, Nelson SD, Sangiorgio SN. Can wear explain the histological variation around metal-on-metal total hips? Clin Orthop Relat Res. 2014;473(2):487–94.
11. Matthies AK, Skinner JA, Osmani H, Henckel J, Hart AJ. Pseudotumors are common in well-positioned low-wearing metal-on-metal hips. Clin Orthop Relat Res. 2012;470(7):1895–906.
12. Reito A, Parkkinen J, Puolakka T, Pajamäki J, Eskelinen A. Diagnostic utility of joint fluid metal ion measurement for histopathological findings in metal-on-metal hip replacements. BMC Musculoskelet Disord. 2015;16(1):393.
13. Grammatopoulos G, Pandit H, Kamali A, Maggiani F, Glyn-Jones S, Gill HS, Murray DW, Athanasou N. The correlation of wear with histological features after failed hip resurfacing arthroplasty. J Bone Joint Surg. 2013;95(12):e81.
14. Langton D, Sidaginamale R, Lord J, Joyce T, Natu S, Nargol A. Metal debris release from taper junctions appears to have a greater clinical impact than debris released from metal on metal bearing surfaces. Bone Joint J. 2013;95(SUPP 1):1–1.
15. Lohmann C, Meyer H, Nuechtern J, Singh G, Junk-Jantsch S, Schmotzer H, Morlock M, Pflüger G. Periprosthetic tissue metal content but not serum metal content predicts the type of tissue response in failed small-diameter metal-on-metal total hip arthroplasties. J Bone Joint Surg. 2013;95(17):1561–8.
16. Nawabi DH, Nassif NA, Do HT, Stoner K, Elpers M, Su EP, Wright T, Potter HG, Padgett DE. What causes unexplained pain in patients with metal-on-metal hip devices? A retrieval, histologic, and imaging analysis. Clin Orthop Relat Res. 2014;472(2):543–54.
17. Paukkeri E, Korhonen R, Hämäläinen M, Pesu M, Eskelinen A, Moilanen T, Moilanen E. The inflammatory phenotype in failed metal-on-metal hip Arthroplasty correlates with blood metal concentrations. PLoS One. 2016; 11(5):e0155121.
18. Liow MHL, Urish KL, Preffer FI, Nielson GP, Kwon Y. Metal ion levels are not correlated with histopathology of adverse local tissue reactions in taper corrosion of total hip arthroplasty. J Arthroplast. 2016;31(8):1797–802.
19. Langton DJ, Jameson SS, Joyce TJ, Gandhi JN, Sidaginamale R, Mereddy P, Lord J, Nargol AV. Accelerating failure rate of the ASR total hip replacement. J Bone Joint Surg Br. 2011;93(8):1011–6.
20. Willert H, Buchhorn GH, Fayyazi A, Flury R, Windler M, Köster G, Lohmann CH. Metal-on-metal bearings and hypersensitivity in patients with artificial hip joints. J Bone Joint Surg. 2005;87(1):28–36.
21. Ricciardi BF, Nocon AA, Jerabek SA, Wilner G, Kaplowitz E, Goldring SR, Purdue PE, Perino G. Histopathological characterization of corrosion product associated adverse local tissue reaction in hip implants: a study of 285 cases. BMC Clin Pathol. 2016;16(1):1.
22. Davies AP, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. J Bone Joint Surg Am. 2005;87(1):18–27.
23. Hasegawa M, Iino T, Sudo A. Immune response in adverse reactions to metal debris following metal-on-metal total hip arthroplasty. BMC Musculoskelet Disord. 2016;17(1):221.
24. Athanasou NA. The pathobiology and pathology of aseptic implant failure. Bone Joint Res. 2016;5(5):162–8.
25. Hart AJ, Satchithananda K, Liddle AD, Sabah SA, McRobbie D, Henckel J, Cobb JP, Skinner JA, Mitchell AW. Pseudotumors in association with well-functioning metal-on-metal hip prostheses: a case-control study using three-dimensional computed tomography and magnetic resonance imaging. J Bone Joint Surg Am. 2012;94(4):317–25.
26. Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. High prevalence of adverse reactions to metal debris in small-headed ASR™ hips. Clin Orthop Relat Res. 2013;471(9):2954–61.
27. Lainiala O, Reito A, Elo P, Pajamäki J, Puolakka T, Eskelinen A. Revision of metal-on-metal hip prostheses results in marked reduction of blood cobalt and chromium ion concentrations. Clin Orthop Relat Res. 2015:1–9.
28. Bills PJ, Racasan R, Underwood R, Cann P, Skinner J, Hart A, Jiang X, Blunt L. Volumetric wear assessment of retrieved metal-on-metal hip prostheses and the impact of measurement uncertainty. Wear. 2012;274:212–9.
29. Searle J, Kerr J, Halliday J, Powell L. Iron storage disease. Pathology Liver. 1994;1:219–41.
30. Scheuer P, Lefkowitch J. Disturbances of copper and iron metabolism. Liver Biopsy Interpretation. 2000. p. 275–83.
31. Petit A, Mwale F, Zukor DJ, Catelas I, Antoniou J, Huk OL. Effect of cobalt and chromium ions on bcl-2, bax, caspase-3, and caspase-8 expression in human U937 macrophages. Biomaterials. 2004;25(11):2013–8.
32. Catelas I, Petit A, Zukor DJ, Huk OL. Cytotoxic and apoptotic effects of cobalt and chromium ions on J774 macrophages–implication of caspase-3 in the apoptotic pathway. J Mater Sci Mater Med. 2001;12(10–12):949–53.
33. Scharf B, Clement CC, Zolla V, Perino G, Yan B, Elci SG, Purdue E, Goldring S, Macaluso F, Cobelli N. Molecular analysis of chromium and cobalt-related toxicity. Sci Rep. 2014;4
34. Catelas I, Petit A, Vali H, Fragiskatos C, Meilleur R, Zukor DJ, Antoniou J, Huk OL. Quantitative analysis of macrophage apoptosis vs. necrosis induced by cobalt and chromium ions in vitro. Biomaterials. 2005;26(15):2441–53.
35. Gallo J, Vaculova J, Goodman SB, Konttinen YT, Thyssen JP. Contributions of human tissue analysis to understanding the mechanisms of loosening and osteolysis in total hip replacement. Acta Biomater. 2014;10(6):2354–66.
36. Perino G, Ricciardi BF, Jerabek SA, Martignoni G, Wilner G, Maass D, Goldring SR, Purdue PE. Implant based differences in adverse local tissue reaction in failed total hip arthroplasties: a morphological and immunohistochemical study. BMC Clin Pathol. 2014;14(1):39.
37. Xia Z, Ricciardi BF, Liu Z, von Ruhland C, Ward M, Lord A, Hughes L, Goldring SR, Purdue E, Murray D. Nano-analyses of wear particles from metal-on-metal and non-metal-on-metal dual modular neck hip arthroplasty. Nanomedicine. 2017;13(3):1205–17.

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2017) 18:523

Page 11 of 11

38. Aloisi F, Pujol-Borrell R. Lymphoid neogenesis in chronic inflammatory diseases. Nat Rev Immunol. 2006;6(3):205–17.
39. Mittal S, Revell M, Barone F, Hardie DL, Matharu GS, Davenport AJ, Martin RA, Grant M, Mosselmans F, Pynsent P. Lymphoid aggregates that resemble tertiary lymphoid organs define a specific pathological subset in metal-on-metal hip replacements. PLoS One. 2013;8(5):e63470.
40. Pandit H, Vlychou M, Whitwell D, Crook D, Luqmani R, Ostlere S, Murray D, Athanasou N. Necrotic granulomatous pseudotumours in bilateral resurfacing hip arthoplasties: evidence for a type IV immune response. Virchows Arch. 2008;453(5):529.
41. Berstock JR, Baker RP, Bannister GC, Case CP. Histology of failed metal-on-metal hip arthroplasty; three distinct sub-types. Hip Int. 2014;24(3):243–8.
42. Phillips EA, Klein GR, Cates HE, Kurtz SM, Steinbeck M. Histological characterization of periprosthetic tissue responses for metal-on-metal hip replacement. J Long Term Eff Med. 2014;24(1):13–23.
43. De Smet K, De Haan R, Calistri A, Campbell P, Ebramzadeh E, Pattyn C, Gill H. Metal ion measurement as a diagnostic tool to identify problems with metal-on-metal hip resurfacing. J Bone Joint Surg. 2008;90(Supplement 4): 202–8.

# PUBLICATION
## II

**Association between periprosthetic tissue metal content, whole blood and synovial fluid metal ion levels and histopathological findings in patients with failed metal-on-metal hip replacement**

Lehtovirta, L., Reito, A., Parkkinen, J., Peräniemi, S., Vepsäläinen, J., & Eskelinen, A.

# Association between periprosthetic tissue metal content, whole blood and synovial fluid metal ion levels and histopathological findings in patients with failed metal-on-metal hip replacement

Lari Lehtovirta[1,2]*, Aleksi Reito[2], Jyrki Parkkinen[3], Sirpa Peräniemi[4], Jouko Vepsäläinen[4], Antti Eskelinen[2]

**1** University of Tampere, Tampere, Finland, **2** Coxa Hospital for Joint Replacement, Tampere, Finland, **3** Fimlab Laboratories, Tampere, Finland, **4** University of Eastern Finland, Kuopio, Finland

\* lehtovirta.lari.a@student.uta.fi

## Abstract

Adverse Reaction to Metal Debris (ARMD) is a major cause of implant failure leading to revision surgery in patients with metal-on-metal (MoM) hip arthroplasties. However, the pathogenesis and its association to implant wear are poorly understood and previous studies have yielded discrepant results. We sought to investigate the associations between histological findings, whole blood and synovial fluid metal ion concentrations and periprosthetic tissue metal concentrations in patients with MoM total hip replacements and hip resurfacings revised for ARMD. 107 hips in total were included in our study. Of these, 87 were total hip replacements and 20 were hip resurfacings, respectively. We found that whole blood, synovial fluid and periprosthetic tissue metal concentrations correlated poorly with histological findings. We suggest that the lack of a clear association between histological findings and wear measures in the present study as well as in previous studies is mostly influenced by variability in patient susceptibility. However, patients presenting with perivascular lymphocytic infiltration had lower chromium concentration in their periprosthetic tissues than patients with no perivascular lymphocytic infiltration. This may reflect the role of metal hypersensitivity in implant failure in these patients. Patients with total hip replacements evinced more necrosis and lymphocytic infiltration in their tissues than patients with hip resurfacings. This suggests that trunnion wear debris is more cytotoxic and/or immunogenic than bearing wear debris leading to higher failure rates seen in patients with total hip replacements.

## Introduction

Adverse Reaction to Metal Debris (ARMD) is a major cause of implant failure leading to revision surgery in patients with metal-on-metal (MoM) hip arthroplasties [1–5]. The term ARMD is an umbrella term describing periprosthetic soft-tissue reactions caused by metal

wear debris that include metallosis, extra-articular pseudotumors (inflammatory, benign soft-tissue masses), overall inflammatory response of the tissue and variable amounts of necrosis seen as caseotic substance on macroscopic level. The term aseptic lymphocytic vasculitis-associated lesions (ALVAL) is more specific and was originally used to describe lymphocytic and necrotic tissue responses [6]. In recent literature terms ALVAL and ARMD have often been inappropriately used interchangeably [7]. The pathogenesis of these adverse reactions is poorly understood, but at least two different mechanisms have been suggested: 1. a non-specific, wear-particle induced cytotoxicity with foreign-body response [8,9] and 2. a specific, type IV hypersensitivity response involving recruitment of lymphocytes in the tissues around failed MoM hip replacements, manifesting as ALVAL [6,10].

Literature regarding implant wear and ARMD is inconclusive. Adverse reactions have been observed both in patients with high- and low-wearing hip replacements [11–15]. Several studies have investigated the associations between wear of the retrieved implants, or indirect markers of wear, and the histopathological findings of periprosthetic tissue taken at the revision surgery, but the results have been discrepant. Lymphocyte-dominated type IV response has been suggested as a cause of failure in patients with low-wearing implants [12,16–18] and cytotoxic response leading to macrophage recruitment in patients with high-wearing implants [8,12,16]. However, conflicting findings not supporting these hypotheses have been published as well [3,15,19–21]. To the best of our knowledge, there has only been one small-scale study that has directly measured the amount of metal in the periprosthetic tissues and analyzed its association with histopathological findings. In that study, Lohmann et. al found that high metal content in the periprosthetic tissue was associated with lymphocyte-dominated, and low metal content was associated with macrophage-dominated response [19]. These findings do not support the hypothesis of metal hypersensitivity as a cause of failure in low-wearing hips and foreign-body cytotoxic response in high-wearing hips.

Studies investigating wear, or indirect measures of wear, and histopathological findings have been inconclusive. The pathogenesis of ARMD and its association to implant wear is poorly understood as well as the potential difference between bearing surface wear debris and taper wear debris in the development of ARMD. Therefore, we aimed to investigate the associations between periprosthetic tissue metal content, whole blood (WB) metal ion concentrations, synovial fluid (SF) metal ion concentrations, and histopathological findings in patients with failed MoM total hip replacements compared to patients with failed MoM hip resurfacings.

## Materials and methods

We recruited a pilot patient for our study in June 2013 followed by the recruitment of consecutive patients between February 2014 and August 2016. In total, 134 hips with MoM implants were revised for ARMD at our institution during the period of recruitment. Of these, two hips were not included due to infection, two hips due to inadequate tissue sample and 23 hips were not included as they were operated on by surgeons who did not participate in recruitment and sample collection. Thus, 107 hips in total were included in our study. Of these, 87 were total hip replacements (THR) and 20 were hip resurfacings, respectively. Whole blood sample was available for 106 patients and synovial fluid sample for 90 patients. In addition to patients undergoing revision surgery, two further patients who had undergone primary hip arthroplasty and whose tissue samples had been retrieved from osteoarthritic synovium were recruited as controls for tissue metal analysis. Surgery was performed by or under the direct supervision of 14 senior orthopedic surgeons. Patient demographics and revised components are presented in detail in Table 1. All patients gave written informed consent to participate in

**Table 1. Patient demographics and implant designs.**

| Patient demographics | | |
|---|---|---|
| Mean age at the time of revision 66.8 years (SD 7.5 years) | | |
| Mean follow-up time between index and revision operation 7.1 years (SD 2.5 years) | | |
| Gender ratio: 42 females (42%) and 57 males (58%) | | |
| **Revised implants** | | |
| **Total hip replacements** | | **Amount** |
| *Femoral component* | *Acetabular component* | |
| DePuy Summit | DePuy ASR | 32 |
| DePuy Summit | DePuy Pinnacle | 10 |
| Biomet Bimetric | Biomet M2A38 | 10 |
| Biomet Bimetric | Biomet ReCap | 5 |
| DePuy Corail | DePuy ASR | 4 |
| Smith-Nephew Synergy | Smith-Nephew R3 | 4 |
| Zimmer ZMR | Zimmer Durom | 2 |
| Zimmer M/L Taper | Zimmer Durom | 2 |
| Wright Medical Profemur | Wright Medical Conserve Plus | 2 |
| Other | Other | 16 |
| | | Total = 87 |
| **Hip resurfacings** | | |
| DePuy ASR | | 9 |
| Smith-Nephew BHR | | 6 |
| Zimmer Durom | | 2 |
| Biomet ReCap | | 2 |
| Smith-Nephew BHR—TM Revision shell | | 1 |
| | | Total = 20 |

https://doi.org/10.1371/journal.pone.0197614.t001

this study, and the study was also approved by the institutional ethical committee (Ethics Committee of Pirkanmaa Hospital District, decision R11196).

Revision surgery was considered, as previously described [2,22–24], if 1) a clear pseudotumour (Imperial class 2A, 2B or 3) [25] was observed on cross-sectional imaging regardless of symptoms or whole blood metal ion levels; or 2) the patient had elevated whole blood metal ion levels and hip symptoms despite normal findings in cross-sectional imaging; or 3) the patient had a continuously symptomatic hip or progressive symptoms regardless of imaging findings or metal ion levels. Symptoms included hip pain, discomfort, sense of instability, and/or impaired function of the hip and sounds from the hip (clacking, squeaking). Whole blood metal ion levels were regarded as being elevated if either chromium or cobalt exceeded 5 ppb. Postoperatively, failure was classified as being due to ARMD and included in our study if the following criteria were met: 1) there was presence of metallosis or macroscopic synovitis in the joint; and/or 2) a pseudotumor was found during revision; and/or 3) a moderate to high number of perivascular lymphocytes along with tissue necrosis and/or fibrin deposition was seen in the histopathologic sample; and 4) perioperatively there was no evidence of component loosening or periprosthetic fracture. In addition, infection was ruled out by obtaining multiple (at least five) bacterial cultures during revision surgery.

## Metal analysis of the periprosthetic tissue

During every revision surgical procedure, samples of the inflamed synovia and/or pseudotumor were obtained for both histopathological and metal content analysis. For metal content analysis, a subsample (approx. 0.3 g) was cut from the tissue sample, weighed, and transferred

into a teflon vessel. Samples were first decomposed with 5 ml suprapur HNO3 (Merck) by microwave digestion technique using a CEM MDS-2000 Microwave System (CEM corporation, Matthews, NC, USA) and then diluted to 10 ml with Milli Q-water. The digests were analyzed for Al, Cr, Co, Ti, Mo, and V with a Inductively Coupled Plasma Optical Emission Spectrometer. Thermo Electron iCAP 6600 Duo View equipped with Cetac ASX-520Hs and autosampler was used (Thermo Fisher Scientific, Waltham, MA, USA). Detection limits for Al, Cr, Co, Ti, Mo, and V were 9.0, 0.2, 0.2, 3.0, 0.2 and 3.0 µg/g, respectively. NIST SRM 1576b (Bovine liver) was used as certified reference material to ensure the performance of analytical procedure for tissue samples.

## Histopathological analysis of the periprosthetic tissue

For histopathological analysis, each tissue sample was formalin fixed and embedded in paraffin. Several 10 µm microtome sections were made. Standard hematoxylin and eosin staining was used. The sections were examined histologically under normal light with a Nikon Eclipse 50i (Nikon Corporation, Shinagawa, Tokyo, Japan). The samples were graded by a senior musculoskeletal pathologist (JP) using grading described by Natu et. al [10]. The grading consisted of following parameters: 1) lymphocyte cuff thickness, 2) whether diffuse lymphocytic infiltration was present, 3) presence of germinal centers, 4) histiocyte sheet thickness, 5) metal particle load within histiocytes, 6) Grade of tissue necrosis, 7) presence of plasma cells and 8) presence of granulomas. Lymphocytic cuff thickness was calculated using a 1mm eyepiece graticule. Calculations were done using 10x magnification. An average of five measurements was taken and graded as 0–3 (absent, 0.25 mm, 0.25–0.75 mm, >0.75 mm). Macrophage sheet thickness was also calculated using a graticule and graded 0–3 (absent, <1 mm, 1–2 mm, >2mm). Metal particle load within macrophages was graded as 0–4 as done in the assessment of iron decomposition in liver cells [26,27]. The extent of overall tissue necrosis in a sample was graded based on the surface necrosis typing according to Davies et al. [28]. Type 1 surface contains intact synovial epithelium. Type 2 surface shows loss of synovial epithelial cells without fibrin deposition. In type 3 surface there is fibrin deposition and in type 4 surface there is extensive necrosis and loss of architecture. The extent of type 4 surface necrosis was used to grade the overall tissue necrosis in a given sample, as described by Natu et al. [10]. In grade 4 necrosis, more than 75% of the tissue sample showed type 4 surface necrosis. In grade 3 necrosis, between 25% and 75% showed type 4 surface necrosis. In grade 2 necrosis either less than 25% of the tissue showed type 4 surface necrosis or the tissue showed type 3 surface. In grade 1 necrosis, the sample consisted of type 2 surface.

## Whole blood and synovial fluid metal analysis

Since January 2012, WB metal ion (Co and Cr) concentrations have been routinely measured as a part of the systematic follow-up program for patients with MoM hip replacements at our institution. All patients underwent WB analysis of Co/Cr following sampling from the antecubital vein using a 21-gauge needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) and trace-element blood tubes containing sodium ethylenediaminetetraacetic acid (EDTA). Standard operating procedures were established at the Finnish Institute for Occupational Health for Co and Cr measurement using dynamic reaction cell inductively coupled plasma (quadripole) mass spectrometry (Agilent 7500 cx, Agilent Technologies, Santa Clara, CA, USA). The laboratory technicians were blinded to all clinical outcomes. The samples were preserved in +6 ˚C to +8 ˚C prior to analysis.

Since October 2011, our MoM hip revision protocol has involved perioperative SF aspiration, which is always taken before opening the deep fascia using a standard 18- to 20-gauge

needle connected to a Vacutainer system (Becton, Dickinson and Company, Franklin Lakes, New Jersey) and trace element tubes containing sodium EDTA. Similar procedures were used for SF metal ion concentration measurement as described above for WB.

## Statistical analysis

Spearman rank correlation was used to study the associations between tissue metal contents, WB and SF metal ion concentrations, and histopathological measures due to these variables being non-normally distributed. Medians were calculated for the tissue metal contents and the histopathological measures. Mann-Whitney U-test was used for comparing medians. When analyzing the correlation between WB metal ion concentrations and other factors, we only included patients with unilateral hip replacements (69 patients with total hip replacement and 9 patients with hip resurfacing) to avoid the confounding effect of metal ions being released to the blood from a second source. The internal validity of our study was investigated by correlating the microscopically visible metal particles with the tissue metal content. We should observe significant association to have a valid method for metal content assessment. The threshold for statistical significance was set to 0.05. The analyses were conducted using IBM SPSS version 21.

## Results

Chromium had the highest concentration of all metals in the periprosthetic tissue in both the HR and THR groups (Table 2). In whole blood, however, cobalt ions were present in higher concentrations than chromium ions (Table 3). Titanium was elevated above the detection limit in nine patients with hip resurfacing and in 29 patients with THR. The concentrations for aluminum and vanadium did not reach the detection limit in any of the patients and were thus omitted from the analyses. There were no statistically significant differences in periprosthetic tissue metal concentrations between THR and hip resurfacing groups (Table 2). In whole blood, median cobalt concentration was approximately twice as high in the THR group compared to hip resurfacing group (Table 3). There was no difference in whole blood chromium ion concentration between the groups (Table 3). In the tissue samples of the two control

**Table 2. Median values with respected p-values and ranges for periprosthetic tissue metal concentrations in patients with total hip replacements (n = 87) and hip resurfacings (n = 20).**

| Metal | Total hip replacement | | Hip resurfacing | | |
|---|---|---|---|---|---|
| | Median concentration in tissue (µg/g) | Range (µg/g) | Median concentration in tissue (µg/g) | Range (µg/g) | P-value |
| Chromium | 39.2 | 0.4–1955.0 | 43.8 | 0.6–922.1 | 0.60 |
| Cobalt | 6.4 | 0.2–262.0 | 3.2 | 0.2–248.8 | 0.189 |
| Molybdenium | 1.8 | 0.2–174.6 | 0.5 | 0.2–32.4 | 0.080 |
| Titanium | 5.8 | 3.0–118.9 | 4.9 | 4.9–25.3 | 0.10 |

https://doi.org/10.1371/journal.pone.0197614.t002

**Table 3. Median values, respected p-values and ranges for whole blood metal ion concentrations in patients with unilateral total hip replacement (n = 69) or hip resurfacing (n = 13) patients.**

| Metal | Total hip replacement | | Hip resurfacing | | |
|---|---|---|---|---|---|
| | Median concentration in whole blood (µg/l) | Range (µg/l) | Median concentration in whole blood (µg/l) | Range (µg/l) | P-value |
| Chromium | 3.7 | 0.4–29.9 | 3.9 | 1.5–7.2 | 0.60 |
| Cobalt | 11.0 | 0.6–108.5 | 3.9 | 1.5–16.2 | 0.001 |

https://doi.org/10.1371/journal.pone.0197614.t003

**Table 4. Lymphocyte cuff thickness, macrophage sheet thickness and grade of necrosis in total hip replacement group (n = 87) and hip resurfacing group (n = 20).**

|  |  | Total hip replacement | Hip resurfacing | P-value |
|---|---|---|---|---|
| Lymphocyte cuff thickness | 0 (absent) | 33 (37.9%) | 15 (75%) |  |
|  | 1 (0–0.25mm) | 41 (47.1%) | 4 (20%) |  |
|  | 2 (>0.25mm) | 13 (14.9%) | 1 (5.0%) |  |
|  |  |  |  | 0.011 |
| Macrophage sheet thickness | 0 (absent) | 1 (1%) | 0 (0%) |  |
|  | 1 (<1mm) | 68 (78.2%) | 18 (90%) |  |
|  | 2 (1-2mm) | 16 (18.4%) | 2 (10%) |  |
|  | 3 (>2mm) | 2 (2.3%) | 0 (0%) | 0.65 |
| Grade of necrosis | 1 | 3 (3.4%) | 8 (40%) |  |
|  | 2 | 19 (21.8%) | 4 (20%) |  |
|  | 3 | 12 (13.8%) | 1 (5%) |  |
|  | 4 | 53 (60.9%) | 7 (35%) |  |
|  |  |  |  | <0.001 |

patients, only the concentration of chromium exceeded the detection limit (0.3 µg/g and 0.5 µg/g, respectively).

Lymphocyte cuff thickness score was higher in patients with THRs versus hip resurfacing (Table 4) and the difference was statistically significant (p = 0.011). Macrophage sheet thickness between hip resurfacing and THR groups did not differ significantly (Table 4, p = 0.65). The grade of tissue necrosis was higher in the THR group compared to hip resurfacing group (Table 4, p < 0.0001).

Correlations between histological variables, periprosthetic tissue metal concentrations, whole blood metal ion levels and synovial fluid metal ion levels are presented in Table 5. Of all the variables, only metal particle load within macrophages had statistically significant but weak correlations with metal ion levels in tissues and whole blood in the THR group. In the resurfacing group, only the synovial fluid chromium and metal particle load had a statistically

**Table 5. Correlations between histological findings, periprosthetic tissue metal concentrations, whole blood metal ion levels (WB) and and synovial fluid (SF) metal ion levels in total hip replacement group (n = 87) and hip resurfacing group (n = 20).** Cells containing statistically significant values are colored in gray.

|  | Total hip replacement | | | | Hip resurfacing | | | |
|---|---|---|---|---|---|---|---|---|
|  | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Metal particle load | Lymphocytic cuffing | Macrophage sheet thickness | Grade of necrosis | Metal particle load |
| Tissue chromium | rho = -0.20 p = 0.063 | rho = 0.022 p = 0.84 | rho = -0.13 p = 0.22 | rho = 0.34 p < 0.01 | rho = -0.36 p = 0.12 | rho = 0.12 p = 0.63 | rho = -0.28 p = 0.23 | rho = 0.29 p = 0.22 |
| Tissue cobalt | rho = -0.072 p = 0.51 | rho = 0.031 p = 0.78 | rho = -0.001 p = 0.99 | rho = 0.30 p < 0.01 | rho = -0.35 p = 0.13 | rho = 0.09 p = 0.72 | rho = -0.28 p = 0.23 | rho = 0.26 p = 0.27 |
| Tissue molybdenium | rho = -0.071 p = 0.514 | rho = 0.060 p = 0.584 | rho = -0.069 p = 0.53 | rho = 0.25 p = 0.019 | rho = -0.29 p = 0.21 | rho = 0.03 p = 0.90 | rho = -0.34 p = 0.15 | rho = 0.30 p = 0.20 |
| Tissue titanium | rho = -0.017 p = 0.88 | rho = -0.036 p = 0.74 | rho = -0.035 p = 0.74 | rho = 0.11 p = 0.30 | rho = -0.16 p = 0.51 | rho = -0.030 p = 0.60 | rho = -0.13 p = 0.58 | rho = -0.077 p = 0.75 |
| WB Cr | rho = -0.092 p = 0.45 | rho = 0.043 p = 0.73 | rho = 0.011 p = 0.92 | rho = 0.21 p = 0.085 | rho = -0.34 p = 0.25 | rho = 0.29 p = 0.35 | rho = 0.14 p = 0.66 | rho = 0.32 p = 0.29 |
| WB Co | rho = -0.088 p = 0.47 | rho = -0.053 p = 0.67 | rho = 0.10 p = 0.41 | rho = 0.39 p < 0.01 | rho = -0.11 p = 0.72 | rho = 0.29 p = 0.35 | rho = 0.33 p = 0.27 | rho = 0.53 p = 0.067 |
| SF Cr | rho = -0.096 p = 0.39 | rho = 0.020 p = 0.86 | rho = -0.077 p = 0.49 | rho = 0.15 p = 0.18 | rho = -0.46 p = 0.22 | rho = 0.00 p = 1.00 | rho = 0.11 p = 0.78 | rho = 0.77 p = 0.016 |
| SF Co | rho = 0.053 p = 0.64 | rho = 0.12 p = 0.30 | rho = 0.17 p = 0.12 | rho = 0.17 p = 0.14 | rho = -0.43 p = 0.25 | rho = 0.21 p = 0.59 | rho = 0.19 p = 0.62 | rho = 0.59 p = 0.096 |

**Table 6. Median metal concentration in tissues with lymphocytes present and tissues with no lymphocytes present in the total hip replacement group (n = 87).**

| Median concentration in tissue (µg/g) | Lymphocytes present | No lymphocytes present | P-value |
|---|---|---|---|
| Chromium | 30.1 | 67.4 | 0.045 |
| Cobalt | 6.4 | 6.1 | 0.43 |
| Molybdenium | 1.7 | 1.8 | 0.38 |

https://doi.org/10.1371/journal.pone.0197614.t006

**Table 7. Median metal concentration in tissues with lymphocytes present and tissues with no lymphocytes present in the hip resurfacing group (n = 20).**

| Median concentration in tissue (µg/g) | Lymphocytes present | No lymphocytes present | P-value |
|---|---|---|---|
| Chromium | 8.0 | 79.3 | 0.11 |
| Cobalt | 1.2 | 4.2 | 0.12 |
| Molybdenium | 0.3 | 0.69 | 0.20 |

https://doi.org/10.1371/journal.pone.0197614.t007

**Table 8. Spearman rho correlation coefficients between tissue metal concentrations, whole blood (WB) and synovial fluid (SF) metal ion concentrations in total hip replacement (n = 87) and hip resurfacing (n = 20) groups.**

| | Total hip replacement | | Hip resurfacing | |
|---|---|---|---|---|
| | *Tissue chromium* | *Tissue cobalt* | *Tissue chromium* | *Tissue cobalt* |
| *WB chromium* | rho = 0.32, p<0.01 | | rho = 0.48, p = 0.10 | |
| *WB cobalt* | | rho = 0.31, p<0.01 | | rho = 0.24, p = 0.43 |
| *SF chromium* | rho = 0.29, p<0.01 | | rho = 0.63, p = 0.067 | |
| *SF cobalt* | | rho = 0.34, p<0.01 | | rho = 0.70, p = 0.035 |

https://doi.org/10.1371/journal.pone.0197614.t008

significant correlation. Correlation between lymphocyte cuff thickness and periprosthetic tissue chromium concentration trended towards significance in the THR group ($\rho$ = -0.20, p = 0.063).

In the THR group in tissues with no lymphocyte infiltration at all, median chromium concentration was higher than in tissues with lymphocyte infiltration present (Table 6). In regard to cobalt and molybdenium there were no statistically significant differences. In the hip resurfacing group, there was a trend towards lower concentrations of chromium and cobalt in those tissues with lymphocytes present but these differences did not reach statistical significance (p = 0.11 and p = 0.12, respectively) (Table 7).

Periprosthetic tissue chromium and cobalt concentrations correlated weakly with whole blood and synovial fluid chromium and cobalt concentrations in THR group (Table 8). In resurfacing group, only synovial fluid cobalt concentration reached statistically significant correlation with periprosthetic tissue cobalt concentration (Table 8).

## Discussion

In the present study, we analyzed periprosthetic tissue metal concentrations, whole blood metal ion concentrations, synovial fluid metal ion concentrations and performed thorough histological analysis of periprosthetic tissue using grading described by Natu et al. [10]. Patients with THR evinced significantly higher amounts of lymphocytes and necrosis in their tissues compared to patients with hip resurfacings despite similar metal concentrations in periprosthetic tissues. Also, patients with total hip replacements had higher whole blood cobalt ion concentrations compared to patients with hip resurfacings. Histological findings that reflect the inflammatory response and necrosis of the tissues correlated poorly with any of the metal

ion measurements. However, periprosthetic tissues with lymphocytic infiltration present had lower amounts of chromium than tissues with no lymphocytic infiltration present.

This study is not without limitations. Firstly, although we performed consecutive recruitment of patients, not all patients who underwent surgery because of ARMD during the recruitment period were included in our study due to some surgeons not participating in the recruitment and some patients being excluded due to infection or inadequate tissue sample. Thus, our series of patients is not completely consecutive. Secondly, we performed semiquantitative histological grading of the samples using grading described by Natu et al. [10]. Grading was done by one observer only. Thirdly, tissue samples used for metal ion measurement were rather small (approx. 0.3g) and may not have completely reflected the average metal concentration of the whole synovium. Also, we were not able to differentiate between metal ions, metals bound to proteins and larger metal particles in the measurement of tissue metal content.

Chromium was the most prominent metal in the periprosthetic tissue in both study groups, which is in line with previous research [19,29–32]. Median concentrations of chromium in the periprosthetic tissue exceeded those of cobalt by more than six-fold in both study groups. On the contrary, in whole blood cobalt ion concentration was higher than that of chromium in the total hip replacement group. Chromium is known to accumulate in the tissues to a high degree while cobalt ions are rapidly transported to the blood and eliminated in the urine [33,34] which explains why chromium concentration is higher than cobalt in periprosthetic tissues and cobalt concentration higher than chromium in whole blood. However, in the hip resurfacing group the cobalt and chromium concentrations in whole blood were similar. This could be due to the small sample size of the hip resurfacing group. Cobalt concentration in whole blood was approximately twice as high in THR group compared to hip resurfacing group while chromium concentrations did not differ between implant groups. Similar findings have been published [35–37]. The excess cobalt in patients with a THR is likely due to material loss at the trunnion surface [38,39]. Periprosthetic metal concentrations correlated poorly with whole blood and synovial fluid metal ion concentrations. The only exception was the good correlation between synovial fluid and periprosthetic tissue cobalt concentrations. We suggest that the overall poor correlations are due to tissues reflecting the accumulated metal load while whole blood and synovial fluid reflect the amount of wear that has been generated more recently. Also, in whole blood and synovial fluid only metal ions are measured whereas in tissues all forms of metal, including particles, ions and metallo-organic complexes, are included in the total amount of metal. Titanium was elevated in 29 patients implying its release from the stem, acetabular cup, or head-neck trunnion. Since this elevation was also seen in patients with hip resurfacings, release from the outer surface of acetabular cup seems probable. Vendittoli et al. found that serum titanium concentrations were indeed higher in hip resurfacings than THRs [40]. In the present study, we did not observe a statistically significant difference in titanium levels between THR and hip resurfacing groups.

We found that periprosthetic tissues retrieved from patients with total hip replacements evinced more severe necrosis and more lymphocytes compared to tissues retrieved from patients with hip resurfacings. Taper wear debris has been suggested to be more immunogenic and cytotoxic than bearing wear debris [41,42]. Xia et al. compared tissues from patients with dual-modular non-MoM implants, MoM THR and MoM hip resurfacings [42]. In dual-modular implants there are two modular junctions which serve as a source of trunnion wear, whereas in THR there is one modular junction and in hip resurfacing there are no modular junctions at all and all wear debris originates from the bearing surfaces. Xia et al. found that tissues from patients with dual-modular implants had highest amounts of lymphocytes and tissue destruction, tissues from THR patients having lower amounts and ultimately tissues from hip resurfacing patients having the lowest amounts. This was despite the fact that tissues from

patients with dual-modular non-mom implants had de facto lowest amount of metal debris. Also, patients with dual-modular implants had shortest time to failure. The authors concluded that trunnion wear is likely more immunogenic and cytotoxic than bearing wear debris, leading to rapid failure. Our results support these findings and suggest that taper wear may cause more tissue destruction than bearing wear manifesting as substantially higher failure rates for THRs than hip resurfacings despite similar amounts of metals in periprosthetic tissues.

In the present study, periprosthetic tissue, whole blood and synovial fluid metal concentrations had poor correlations with histological findings. Several retrieval studies have been conducted to study the relationship between implant wear and histopathological findings. Campbell et al. investigated the amount of implant wear and type of tissue response in patients with failed MoM hips and found that low wear was associated with a hypersensitivity type lymphocytic response [12]. Conversely, high component wear was associated with a macrophage-dominated response suggesting non-specific wear-related cytotoxicity. Slightly differently, Grammatopoulos et al. found that implant wear was associated with the number of macrophages but not with the number of lymphocytes [8]. In their study, all patients with a pseudo-tumor and a low-wearing implant had a high ALVAL score suggesting a hypersensitivity response. However, most pseudotumors were associated with highly worn prostheses. A recent study by Paukkeri et al. found that whole blood chromium and cobalt ion correlations, indirect markers of wear, were higher in patients with macrophage-dominated response and lower in patients with lymphocyte-dominated response. On the contrary, Liow et al. found no correlation between whole blood metal ion levels and histological findings in periprosthetic tissue. To the best of our knowledge, only one previous study has investigated the periprosthetic metal content in relation to histopathological findings in patients with failed MoM hip arthroplasties [19]. Lohmann et al. found that high periprosthetic tissue metal content (chromium, cobalt and nickel combined and separately) was associated with a lymphocyte-dominated response and low metal content with a macrophage-dominated response. We would like to address some weaknesses in the study which may have affected the outcome. Firstly, the small number of cases in that study is likely to be a limiting factor. There were only five patients in the macrophage-dominated group and 22 patients in the lymphocyte-dominated group. The high incidence for the lymphocyte-dominated response compared to the macrophage-dominated response is neither supported by previous studies [8,10,12] nor the results of our study. Furthermore, mean values for tissue metal concentration were calculated and compared between the two groups. With nonparametric variables, this is not a valid statistical method. In conclusion, literature regarding the association between histological findings and wear or indirect measures of wear is very discrepant. A recent review suggested that periprosthetic tissue metal concentrations may correlate more accurately with the histology than serum metal ion levels [7]. Our results do not support that hypothesis. We found that tissue metal concentrations as well as whole blood and synovial fluid metal ion concentrations had poor correlations with histological findings. However, tissues with lymphocytic infiltration had lower amounts of chromium compared to tissues with no lymphocytic infiltration. This finding alone supports the hypothesis of hypersensitivity as a cause of failure in patients with low-wearing MoM hip implants. However, there was no correlation between the amount of lymphocytes and periprosthetic chromium concentration, which makes it difficult to draw conclusions in light of the overall results.

Associations between histological findings and wear or indirect measures of wear has been inconsistent and weak in previous studies as well as the present study. In the literature, the histopathology of ARMD tissues has mainly been categorized into a wear-related foreign-body response or a supposedly hypersensitivity-related lymphocyte-type response or a mix of both. It is possible and probable that some patients have both high-wearing implants and an

underlying hypersensitivity-type response that would have evoked even in the presence of a low-wearing implant. This combination may result in a mixed-type tissue response that has the characteristics of both wear-related innate immune responses and hypersensitivity-related adaptive tissue responses and therefore makes it difficult to distinguish between the two based on the tissue metal content or some other measure of wear. This may also explain why we did not find correlation between tissue metal concentrations and lymphocytes, but did find a difference in concentration of metals between those with no lymphocytes versus those with lymphocytes present. It is possible that the differences between different lymphocyte scores are too subtle and vulnerable to error for a statistically significant correlation to be detected between these scores and metal concentrations in tissues. In contrast, dividing the patients in two groups: those with perivascular lymphocytes and those without, may thus reflect the association between inflammatory response and tissue metal concentration more clearly. Also, trunnion wear from THR appears to elicit different tissue responses than bearing wear, which makes comparison between studies difficult. In numerous previous studies, patient susceptibility has been suggested as an important factor contributing to the development of ARMD [12,14,15,17]. Patient susceptibility means that patients can elicit different types of responses to the metal debris at different levels of metal load in their tissues. We suggest that variability in the threshold level of metal debris needed to cause significant tissue responses explains the weakness and inconsistency between histological findings and wear measurements. The role of patient susceptibility in the pathogenesis of ARMD warrants further research.

## Conclusions

In conclusion, periprosthetic tissue metal concentrations had poor correlation with histological findings or metal ion levels in whole blood and synovial fluid. We suggest that this is mostly due to variation in patient susceptibility manifesting as individually different levels of reactivity to metal debris. Despite the similar metal concentrations in periprosthetic tissues, patients with THR evinced more lymphocytes and necrosis in their tissues compared to patients with hip resurfacings. We suggest that taper wear debris from THR is more immunogenic or cytotoxic compared to bearing surface wear debris, leading to higher failure rates in patients with THRs compared to hip resurfacings. In THR, tissues with lymphocytic infiltration had lower amounts of chromium than tissues with no lymphocytic infiltration. Similar trend was observed in hip resurfacings, but this did not reach statistical significance. These findings alone support the hypothesis of metal hypersensitivity as a cause of failure in a subgroup of patients with low-wearing hip implants. Interestingly, however, we did not observe correlation between lymphocyte scores and periprosthetic tissue chromium concentrations. Thus, it is difficult to draw solid conclusions regarding the role of metal hypersensitivity as a cause of failure in patients with low-wearing hip implants.

## Acknowledgments

## Author Contributions

**Conceptualization:** Lari Lehtovirta, Aleksi Reito, Antti Eskelinen.

**Data curation:** Lari Lehtovirta, Aleksi Reito.

**Formal analysis:** Lari Lehtovirta.

## References

1. Barrett WP, Kindsfater KA, Lesko JP. Large-diameter modular metal-on-metal total hip arthroplasty: incidence of revision for adverse reaction to metallic debris. J Arthroplasty. 2012; 27: 976–983. e1. https://doi.org/10.1016/j.arth.2012.01.019 PMID: 22425300

2. Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. Outcome of Birmingham hip resurfacing at ten years: role of routine whole blood metal ion measurements in screening for pseudotumours. Int Orthop. 2014; 38: 2251–2257. https://doi.org/10.1007/s00264-014-2429-4 PMID: 25030963

3. Langton DJ, Jameson SS, Joyce TJ, Gandhi JN, Sidaginamale R, Mereddy P, et al. Accelerating failure rate of the ASR total hip replacement. J Bone Joint Surg Br. 2011; 93: 1011–1016. https://doi.org/10.1302/0301-620X.93B8.26040 PMID: 21768621

4. Reito A, Lainiala O, Elo P, Eskelinen A. Prevalence of failure due to adverse reaction to metal debris in modern, medium and large diameter metal-on-metal hip replacements–the effect of novel screening methods: systematic review and metaregression analysis. PloS one. 2016; 11: e0147872. https://doi.org/10.1371/journal.pone.0147872 PMID: 26930057

5. Bosker B, Ettema H, Boomsma M, Kollen B, Maas M, Verheyen C. High incidence of pseudotumour formation after large-diameter metal-on-metal total hip replacement A prospective cohort study. Journal of Bone & Joint Surgery, British Volume. 2012; 94: 755–761.

6. Willert H, Buchhorn GH, Fayyazi A, Flury R, Windler M, Köster G, et al. Metal-on-Metal Bearings and Hypersensitivity in Patients with Artificial Hip Joints. The Journal of Bone & Joint Surgery. 2005; 87: 28–36.

7. Athanasou NA. The pathobiology and pathology of aseptic implant failure. Bone Joint Res. 2016; 5: 162–168. https://doi.org/10.1302/2046-3758.55.BJR-2016-0086 PMID: 27146314

8. Grammatopoulos G, Pandit H, Kamali A, Maggiani F, Glyn-Jones S, Gill HS, et al. The correlation of wear with histological features after failed hip resurfacing arthroplasty. The Journal of Bone & Joint Surgery. 2013; 95: e81.

9. Mahendra G, Pandit H, Kliskey K, Murray D, Gill HS, Athanasou N. Necrotic and inflammatory changes in metal-on-metal resurfacing hip arthroplasties: relation to implant failure and pseudotumor formation. Acta orthopaedica. 2009; 80: 653–659. https://doi.org/10.3109/17453670903473016 PMID: 19995315

10. Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol AV. Adverse reactions to metal debris: histopathological features of periprosthetic soft tissue reactions seen in association with failed metal on metal hip arthroplasties. J Clin Pathol. 2012; 65: 409–418. https://doi.org/10.1136/jclinpath-2011-200398 PMID: 22422805

11. Langton D, Joyce T, Jameson S, Lord J, Van Orsouw M, Holland J, et al. Adverse reaction to metal debris following hip resurfacing the influence of component type, orientation and volumetric wear. Journal of Bone & Joint Surgery, British Volume. 2011; 93: 164–171.

12. Campbell P, Ebramzadeh E, Nelson S, Takamura K, De Smet K, Amstutz H. Histological Features of Pseudotumor-like Tissues From Metal-on-Metal Hips. Clinical Orthopaedics and Related Research®. 2010; 468: 2321–2327.

13. Kwon Y, Ostlere SJ, McLardy-Smith P, Athanasou NA, Gill HS, Murray DW. "Asymptomatic" pseudotumors after metal-on-metal hip resurfacing arthroplasty: prevalence and metal ion study. J Arthroplasty. 2011; 26: 511–518. https://doi.org/10.1016/j.arth.2010.05.030 PMID: 20591612

14. Matthies AK, Skinner JA, Osmani H, Henckel J, Hart AJ. Pseudotumors are common in well-positioned low-wearing metal-on-metal hips. Clinical Orthopaedics and Related Research®. 2012; 470: 1895–1906.

15. Ebramzadeh E, Campbell P, Tan TL, Nelson SD, Sangiorgio SN. Can wear explain the histological variation around metal-on-metal total hips? Clinical Orthopaedics and Related Research®. 2014; 473: 487–494.

16. Paukkeri E, Korhonen R, Hämäläinen M, Pesu M, Eskelinen A, Moilanen T, et al. The Inflammatory Phenotype in Failed Metal-On-Metal Hip Arthroplasty Correlates with Blood Metal Concentrations. PloS one. 2016; 11: e0155121. https://doi.org/10.1371/journal.pone.0155121 PMID: 27227536

17. Ebramzadeh E, Campbell PA, Takamura KM, Lu Z, Sangiorgio SN, Kalma JJ, et al. Failure modes of 433 metal-on-metal hip implants: how, why, and wear. Orthop Clin North Am. 2011; 42: 241–250. https://doi.org/10.1016/j.ocl.2011.01.001 PMID: 21435498

18. Nawabi DH, Nassif NA, Do HT, Stoner K, Elpers M, Su EP, et al. What causes unexplained pain in patients with metal-on metal hip devices? A retrieval, histologic, and imaging analysis. Clinical Orthopaedics and Related Research®. 2014; 472: 543–554.

19. Lohmann C, Meyer H, Nuechtern J, Singh G, Junk-Jantsch S, Schmotzer H, et al. Periprosthetic tissue metal content but not serum metal content predicts the type of tissue response in failed small-diameter metal-on-metal total hip arthroplasties. The Journal of Bone & Joint Surgery. 2013; 95: 1561–1568.

20. Reito A, Parkkinen J, Puolakka T, Pajamäki J, Eskelinen A. Diagnostic utility of joint fluid metal ion measurement for histopathological findings in metal-on-metal hip replacements. BMC musculoskeletal disorders. 2015; 16: 1.

21. Liow MHL, Urish KL, Preffer FI, Nielson GP, Kwon Y. Metal ion levels are not correlated with histopathology of adverse local tissue reactions in taper corrosion of total hip arthroplasty. J Arthroplasty. 2016; 31: 1797–1802. https://doi.org/10.1016/j.arth.2016.01.041 PMID: 26923497

22. Reito A, Moilanen T, Puolakka T, Pajamäki J, Eskelinen A. Repeated metal ion measurements in patients with high risk metal-on-metal hip replacement. Int Orthop. 2014; 38: 1353–1361. https://doi.org/10.1007/s00264-014-2300-7 PMID: 24638214

23. Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. High prevalence of adverse reactions to metal debris in small-headed ASR™ hips. Clinical Orthopaedics and Related Research®. 2013; 471: 2954–2961.

24. Reito A, Elo P, Puolakka T, Pajamäki J, Eskelinen A. Femoral diameter and stem type are independent risk factors for ARMD in the large-headed ASR THR group. BMC musculoskeletal disorders. 2015; 16: 1.

25. Hart AJ, Satchithananda K, Liddle AD, Sabah SA, McRobbie D, Henckel J, et al. Pseudotumors in association with well-functioning metal-on-metal hip prostheses: a case-control study using three-dimensional computed tomography and magnetic resonance imaging. J Bone Joint Surg Am. 2012; 94: 317–325. https://doi.org/10.2106/JBJS.J.01508 PMID: 22336970

26. Searle J, Kerr J, Halliday J, Powell L. Iron storage disease. Pathology of the liver. 1994; 1: 219–241.

27. Scheuer P, Lefkowitch J. Disturbances of copper and iron metabolism. Liver biopsy interpretation. 1980: 189–192.

28. Davies AP, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. J Bone Joint Surg Am. 2005; 87: 18–27. https://doi.org/10.2106/JBJS.C.00949 PMID: 15634811

29. Catelas I, Medley JB, Campbell PA, Huk OL, Bobyn JD. Comparison of in vitro with in vivo characteristics of wear particles from metal–metal hip implants. Journal of Biomedical Materials Research Part B: Applied Biomaterials. 2004; 70: 167–178.

30. Catelas I, Campbell PA, Bobyn JD, Medley JB, Huk OL. Wear particles from metal-on-metal total hip replacements: effects of implant design and implantation time. Proc Inst Mech Eng H. 2006; 220: 195–208. https://doi.org/10.1243/09544119JEIM112 PMID: 16669387

31. Hart AJ, Quinn PD, Sampson B, Sandison A, Atkinson KD, Skinner JA, et al. The chemical form of metallic debris in tissues surrounding metal-on-metal hips with unexplained failure. Acta biomaterialia. 2010; 6: 4439–4446. https://doi.org/10.1016/j.actbio.2010.06.006 PMID: 20541630

32. Scharf B, Clement CC, Zolla V, Perino G, Yan B, Elci SG, et al. Molecular analysis of chromium and cobalt-related toxicity. Scientific reports. 2014; 4.

33. Merritt K, Crowe TD, Brown SA. Elimination of nickel, cobalt, and chromium following repeated injections of high dose metal salts. J Biomed Mater Res. 1989; 23: 845–862. https://doi.org/10.1002/jbm.820230804 PMID: 2777829

34. Brown SA, Zhang K, Merritt K, Payer JH. In vivo transport and excretion of corrosion products from accelerated anodic corrosion of porous coated F75 alloy. J Biomed Mater Res. 1993; 27: 1007–1017. https://doi.org/10.1002/jbm.820270806 PMID: 8408113

35. Hart AJ, Sabah SA, Sampson B, Skinner JA, Powell JJ, Palla L, et al. Surveillance of Patients with Metal-on-Metal Hip Resurfacing and Total Hip Prostheses: A Prospective Cohort Study to Investigate the Relationship Between Blood Metal Ion Levels and Implant Failure. J Bone Joint Surg Am. 2014; 96: 1091–1099. https://doi.org/10.2106/JBJS.M.00957 PMID: 24990974

36. Lainiala OS, Moilanen TP, Hart AJ, Huhtala HS, Sabah SA, Eskelinen AP. Higher Blood Cobalt and Chromium Levels in Patients With Unilateral Metal-on-Metal Total Hip Arthroplasties Compared to Hip Resurfacings. J Arthroplasty. 2016; 31: 1261–1266. https://doi.org/10.1016/j.arth.2015.11.045 PMID: 26775067

37. Renner L, Faschingbauer M, Schmidt-Braekling T, Boettner F. Cobalt serum levels differ in well functioning Birmingham resurfacing and Birmingham modular THA. Arch Orthop Trauma Surg. 2016; 136: 715–721. https://doi.org/10.1007/s00402-016-2439-3 PMID: 26983720

38. Matthies AK, Racasan R, Bills P, Blunt L, Cro S, Panagiotidou A, et al. Material loss at the taper junction of retrieved large head metal-on-metal total hip replacements. Journal of Orthopaedic Research. 2013; 31: 1677–1685. https://doi.org/10.1002/jor.22431 PMID: 23918742

39. Hothi HS, Berber R, Whittaker RK, Blunn GW, Skinner JA, Hart AJ. The Relationship Between Cobalt/Chromium Ratios and the High Prevalence of Head-Stem Junction Corrosion in Metal-on-Metal Total Hip Arthroplasty. J Arthroplasty. 2016; 31: 1123–1127. https://doi.org/10.1016/j.arth.2015.11.014 PMID: 26796774

40. Vendittoli P, Roy A, Mottard S, Girard J, Lusignan D, Lavigne M. Metal ion release from bearing wear and corrosion with 28 mm and large-diameter metal-on-metal bearing articulations A FOLLOW-UP STUDY. Journal of Bone & Joint Surgery, British Volume. 2010; 92: 12–19.

41. Langton D, Sidaginamale R, Lord J, Joyce T, Natu S, Nargol A. Metal debris release from taper junctions appears to have a greater clinical impact than debris released from metal on metal bearing surfaces. Bone Joint J. 2013; 95: 1–1.

42. Xia Z, Ricciardi BF, Liu Z, von Ruhland C, Ward M, Lord A, et al. Nano-analyses of wear particles from metal-on-metal and non-metal-on-metal dual modular neck hip arthroplasty. Nanomedicine: Nanotechnology, Biology and Medicine. 2017; 13: 1205–1217.

# PUBLICATION
## III

**Histopathological patterns seen around failed metal-on-metal hip replacements: Cluster and latent class analysis of patterns of failure.**

Reito, A., Lehtovirta, L., Parkkinen, J., & Eskelinen, A.

1  # Histopathological patterns seen around failed metal-on-metal hip

2  # replacements: a cluster and a latent class analysis of patterns of

3  # failure

4  *Running title: Histopathology and failed metal-on-metal hips*

5  **Aleksi Reito[1], MD, PhD**

6  Lari Lehtovirta[1,3], BM

7  Jyrki Parkkinen[1,2] MD, PhD

8  Antti Eskelinen[1], MD, PhD

9

10  **Affiliations**:

11  [1]Coxa Hospital for Joint Replacement, Tampere, Finland

12  [2]Fimlab Laboratories, Tampere, Finland

13  [3]School of Medicine, Tampere University, Tampere, Finland

14

15  **Corresponding author**:

16  Aleksi Reito

17  Biokatu 6b

18  33520 Tampere

19  Finland

20  aleksi@reito.fi

21  tel. +358 14 269 3168

22  fax. +358 3 117 8090

23
24

25  **Abstract**

26  We aimed to establish latent subtypes of histopathological patterns in failed metal-on-metal hip

27  replacements. Tissue samples of the synovia from the neocapsule were retrieved from 284

28  revised ASR hip replacements and analyzed histologically. Hierarchical cluster analysis and

29  polytomous latent class analysis were performed to establish the underlying structure and

30  relationships of the histological observations and to find similar cohorts of cases. Clustering

31  analyses suggested four distinct subtypes that could be readily and reasonably labeled and

32  mapped against a recent consensus statement. The first two subtypes showed synovial necrosis,

33  lymphocyte sheets and abundant or thin histiocyte sheets. In addition, the first subtype showed

34  abundant germinal centers and no metal particles either extra or intracellularly. Metal particles

35  were, however, seen in the second subtype. Hence, the first subtype was labeled "immunologic

36  type IV neosynovitis" and the second subtype "abrasion induced inflammatory lymphocytic type

37  I neosynovitis". The third and fourth subtypes showed no perivascular and diffuse lymphocytes,

38  but a higher number of metal particles intra and extracellularly. The third subtype had synovial

39  necrosis along with granulomas and was labeled "abrasion induced necrotic Type I

40  neosynovitis", whereas the fourth subtype had readily intact synovial lining, and this subtype

41  was labeled "abrasion-induced foreign body type I neosynovitis". Histopathological findings in

42  failed MoM hips are not just one wide entity. These hips evince four different histological

43  patterns that also differ at the macroscopic level. Moreover, the often stated "ALVAL-type

44  reaction" seems to be dualistic in nature, which is a novel finding.

45  **Keywords**: metal-on-metal, hip replacement, histology, neosynovitis, ALVAL

46

47

48

**Introduction**

Soft tissue reactions in traditional non-metal-on-metal (MoM) hip replacements are very rare and most reoperations are due to other reasons, such as aseptic loosening, dislocation, osteolysis, hardware failure, and infection. In patients operated on with MoM hip replacements, adverse reaction to metal debris (ARMD) has become the most common reason for revision surgery[1]–[3]. The manifestation of the adverse soft tissue reactions seen in failed MoM hip replacements is very variable both at the macroscopic and the microscopic level. Moreover, there is no universal consensus as to what constitutes ARMD.

Histological findings in failed MoM hip replacements are variable. Aseptic lymphocyte-vasculitis associated lesion (ALVAL) reaction can be seen in failed MoM hip replacements [4]. Since then "ALVAL-type reaction" has been used as an umbrella term for a specific spectrum of histological findings. However, no robust attempts have been made to subdivide this histological reaction. Cases with ALVAL-reaction include infiltration of perivascular T-lymphocytes in postcapillary venules, synovial destruction and fibrin exudation, and variable numbers of diffuse lymphocytes [4]–[6]. Macrophage-dominated foreign body reaction is also commonly seen in failed MoM hip replacements [7]–[9] . Related to this conceptually dualistic nature of histopathology, a 10-point ALVAL-score system has been proposed to distinguish between wear-related and hypersensitivity-related failures in MoM hip replacements [6]. Some authors have, however, suggested an additional histopathological subtype that involves synovial necrosis with macrophage-dominated inflammatory response without notable perivascular lymphocyte infiltration [9],[10]. Further contradicting the dualistic behavior, other histological patterns have

71  been seen in subsets of patients with failed MoM hip replacements that include sarcoid-like

72  granulomas and lymphoid aggregates (7),(11).

73  Due to the extreme versatility seen at both the macroscopic and microscopic level in patients

74  with failed MoM hip replacements, it is very unlikely that these adverse soft tissue reactions are

75  the result of only one or two different cascades of events starting from molecular interactions

76  that may lead to severe macroscopic soft tissues changes (12). It is important therefore to

77  understand all the possible subtypes of histopathological responses in failing MoM hips because

78  patients with different types of responses may have a very different prognosis and may require a

79  different type of follow-up.

80

81  The purpose of our study was to investigate the underlying patterns of histopathological findings

82  seen in failed MoM hip replacements and to ascertain possible histopathological subtypes

83  according to the histopathological consensus classification of joint implant related pathology (13)

84  . To achieve this aim, we performed an exploratory cluster and latent class analysis to identify

85  possible clusters of cases and observations suggestive of different subtypes of histopathological

86  findings seen in the synovia of patients with failed MoM hip replacements.

87

88  **Materials and Methods**

89  *Study population*

90  One thousand and thirty-six ASR MOM hip replacements were performed on 887 patients at our

91  institution between March 2004 and December 2009. All living and non-revised patients were

92  subjected to a screening protocol starting in September 2010. The aim of the screening protocol

93  was to detect articulation-related soft tissue reactions in these patients. All patients were referred

94     for cross-sectional imaging and whole blood chrome and cobalt ion measurement. Imaging

95     findings were graded using the Imperial classification (14).

96     *Revisions and revision procedures*

97     As of January 2015, a total of 334 hips in 301 patients had undergone revision surgery since the

98     beginning of the screening. In 296 of the 334 revision surgeries, a tissue sample was retrieved for

99     histopathological analysis ARMD being the most common reason for revision surgery (Figure

100     1). The diagnosis of ARMD was based on pre- and perioperative findings as previously

101     described (15),(16). In each revision surgery performed due to non-infectious reasons, 1 to 5

102     samples of synovial and/or pseudotumor (PT) tissues were routinely collected for histological

103     analysis. 99 hips had evidence of PT prior to revision surgery. The primary site of the sample

104     was the PT capsule (ie. PT sac or bursae wall), and if such tissue was not present, a sample was

105     taken from the pseudocapsule. A synovial fluid sample was also taken in order to assess the

106     chrome and cobalt ion levels in the fluid.

107     *Histology*

108     Each tissue sample was formalin-fixed and embedded in paraffin. Several 10 µm microtome

109     sections were made. Standard hematoxylin and eosin staining was used. The sections were

110     examined histologically under normal light Nikon Eclipse 50i microscope (Nikon Corporation,

111     Shinagawa, Tokyo, Japan). For investigational purposes, all available samples were

112     retrospectively analyzed by a pathologist with more than 10 years of experience in the field (JP)

113     according to principles described by Natu et al. (17). Necrosis was classified from Grade I to IV

114     according to Natu et al. Natu grading of necrosis is based on loss in synovial integrity described

115     by Davies et al. (5). Davies Type 1 synovial surface is intact epithelium. Type 2 synovial surface

116     has loss of cellular lining but without fibrin exudation. Type 3 synovial surface has both loss of

117     synovial cell lining and fibrin exudate is present. Type 4 synovial surface has gross disruption,

118     fissuring, and fibrin exudates. In Natu Grade I necrosis, the synovial surface consisted of only

119     Davies Type 1 or 2. Grade II consisted of Davies Type 3 or maximally 25% of Type 4. In Grade

120     III necrosis, the surface consisted of between 25% and 75% of Davies Type IV. In Grade IV, the

121     surface showed more than 75% of Davies Type IV synovial loss.  Lymphocytic cuff thickness

122     was calculated using a 1 mm eyepiece graticule. Calculations were done using 10x

123     magnification. An average of five measurements were taken and graded as 0 to 3 (absent, 0.25

124     mm, 0.25 mm to 0.75 mm, >0.75 mm). The thickness of histiocyte sheets was calculated and

125     graded 0 to 3 (absent, <1 mm, 1 mm to 2 mm, >2mm). Particle load within histiocytes was

126     graded as used in the assessment of iron decomposition in liver cells as described by Natu et al.

127     [17].  Extracellular metal particles were defined to be present if they were visible with only x10

128     magnification or with the naked eye in the section. Diffuse synovitis was defined as described by

129     Natu et al. Similarly, the presence of germinal centers was noted. Plasma cell forming aggregates

130     were also assessed.

131     *Statistics*

132     Two different cluster-based segmentation methods were used to detect the underlying latent

133     groupings of cases. Latent class analysis (LCA) and cluster analysis with hierarchical approach

134     (HCA) were used [18],[19]. Between-group differences in clinical variables after HCA were

135     compared using either Kruskal-Wallis -test or Chi-squared test.

136     *Cluster analysis*

137     Our main interest was to find clusters of cases (hips) based on their dissimilarity. Since our data

138     comprised binary and ordinal variables, we chose the Gower method to form the distance matrix

139     [20]. After the selection of the appropriate dissimilarity measurement, clustering began by

140   assigning each case to be an individual cluster forming a proximity matrix sized 284x284. The

141   matrix reflected the closeness of each cluster. Each case began as an individual cluster and was

142   gradually merged with the most closely related cluster (of cases). We used the complete linkage

143   method. This process was repeated until one single cluster remained. Our aim was to identify any

144   meaningful and histologically relevant clusters. Hence, we did not use solely the agglomerative

145   approach, which is the most commonly used method, to establish the optimal number of clusters.

146   The agglomerative process uses the agglomeration schedule in which the change in

147   agglomeration coefficient is depicted as the distance between merged clusters. The higher the

148   change in the agglomeration coefficient the higher is the dissimilarity between clusters. We

149   interpreted the last stages of the clustering process to define the meaningful clusters of

150   observations since five or less clusters were expected to be seen. "Natural break" was defined as

151   the largest change in agglomeration coefficient producing meaningfully distinguishable clusters.

152   *Latent class analysis*

153   Latent class analysis (LCA) was also performed to further analyze the possible underlying

154   structures in our data set. LCA also aims to identify meaningful groups or class memberships of

155   cases according to their (dis)similarity. To identifying the optimal set of groups, latent class

156   analysis was first performed with two groups, then three groups, and so on. Akaikes Information

157   Criterion (AIC) indices were interpreted to assess the most suitable baseline model. As with

158   cluster analysis, with latent class analysis we aimed to have both the optimal model suggested by

159   the indices and to have a meaningful set of groups so that each group could be readily labeled.

160   *Validation analysis*

161   We aimed to validate our primary outcome after cluster analysis and LCA by first running a

162   validation analysis using both techniques and then separately for both implant groups. The

163 rationale for this was the different wear behaviors between stemmed total hip arthroplasties

164 (THA) and hip resurfacings (HR). Bearing wear is seen in both implants, but taper corrosion is

165 only seen in THA (21). If our segmentation techniques are robust against one major etiological

166 factor, similar clusters and class memberships should be produced regardless of the implant type

167 included in the analysis.

168

169 Each cluster formed by validation clustering was matched against primary clusters. The

170 distribution of cases among clusters obtained from validation cluster analysis was cross-tabulated

171 against primary clustering to see whether discorcondant cases, i.e., negative matches among two

172 different clustering processes, existed. Validation LCA was performed in an equal way using the

173 same principle as with the study cohort (all cases included).

174

175 **Results**

176 Revision surgery was performed on 296 hips from which a tissue sample was retrieved for

177 analysis. In twelve cases, the sample was destroyed or contained no viable tissue and was

178 excluded. The final study cohort included 284 tissue samples. The distribution of histological

179 findings divided by implant type is shown in Table 1.

180 *Number of subgroups*

181 In the HCA, the change in agglomeration coefficient and the value of the AIC in the LCA were

182 interpreted to evaluate the number of meaningful subgroups (Figure 2). The highest change in

183 agglomeration coefficient was seen if two subgroups were selected. This contradicted our pre-

184 study assumption, supported by the previous literature, of having >2 subgroups. The second

185 highest change in the agglomeration coefficient was seen with the formation of four different

186 subgroups. In the LCA, the AIC was lowest with both a four- and five-group solution. The

187 difference was, however, minimal. The five-group solution was not supported in the HCA based

188 on the change in the agglomeration coefficient. We considered the four-group solution to be the

189 most informative, i.e., four clusters in HCA and four classes in LCA.

190 *Characteristics and labeling of the subgroups*

191 HCA and LCA resulted in four similar and meaningful subgroups (Table 2 and 3). Cluster 1 in

192 HCA and Class 1 in LCA could be readily labeled as "abrasion-induced foreign body type I

193 neosynovitis". The characteristics of this subgroup were absent or mild necrosis, lack of diffuse

194 synovitis, germinal centers, plasma cells, and granulomas. Perivascular lymphocyte cuffs were

195 also mainly absent. Histiocyte sheets were mainly thin and particle load inside them was

196 moderate. Cluster 2 in HCA and Class 2 in LCA were labeled as "abrasion induced necrotic

197 Type I neosynovitis" (Figure 3). As in the foreign body reaction, diffuse synovitis, plasma cells,

198 germinal centers, and perivascular lymphocytic cuffs were absent. However, granulomas were

199 seen; histiocyte sheets were thicker; the level of necrosis was moderate; the particle load inside

200 histiocytes was higher, and extracellular metals particles were present. Cluster 3 in HCA and

201 Class 3 in LCA were similar. Plasma cells and germinal centers were prevalent; the level of

202 necrosis was very high; lymphocytic cuffs were thick, and both particle load and extracellular

203 metal content was low or absent. We labeled Cluster 3 in HCA and Class 3 in LCA as

204 "immunologic type IV neosynovitis" (Figure 4). The remaining subgroups, i.e., Cluster 4 in

205 HCA and Class 4 in LCA, were partly similar to the previous reaction, but plasma cells and

206 germinal centers were less frequent and lymphocytic cuffs were thinner. Necrosis was, however,

207 moderate or high and also diffuse synovitis was most frequent. Extracellular metal particles were

208 commonly present and particle load within histiocytes was high, and hence these subgroups were

209 labeled as "abrasion induced inflammatory lymphocytic type I neosynovitis".

210

211 The hips in the "immunologic neosynovitis" group were prominently hip resurfacings and they

212 had the highest median level of cobalt ions in WB and synovial fluid (Table 4). Thick-walled

213 PTs were most common in the "immunologic neosynovitis" and "inflammatory lymphocytic

214 neosynovitis" groups. The lowest cobalt ion levels were seen in the "foreign body reaction"

215 group.

216 *Interpretation of other subgroups in HCA*

217 HCA with three clusters indicated that clusters 1 and 3, i.e., the foreign body and cytotoxic

218 reaction subgroups, were merged. Since these clusters evince very distinct characteristics, a

219 three- subgroup solution was not deemed meaningful, as earlier described.

220 *Validation analysis*

221 In THA group, the HCA and LCA were done using four clusters and four classes. The

222 characteristics of the these clusters and classes were similar to those when clustering the whole

223 cohort. Thus, the same labeling was meaningful (see supplement A).  In HCA, a positive match

224 regarding distribution of cases among clusters was seen in 143 of 206 cases (69.5%.) (Table 5).

225 A mismatch was therefore seen in 75 hips, and the majority of these were due to 39 hips with a

226 primary analysis that suggested "immunologic neosynovitis" group, but a sensitivity analysis that

227 suggested "inflammatory lymphocytic neosynovitis" group. In the LCA validation analysis for

228 THAs, a match was seen in 163 of 206 (79.1%) cases (Table 6). Due to the small absolute

229 number of cases in the histological counts, validation analysis was not performed for HRs**.**

230

**Discussion**

231

232  The results of the hierarchical cluster and latent class analysis in the current study implied four

233  distinct subtypes of histopathological findings in failed ASR MoM hip replacements (Table 7).

234  We propose that in addition to traditional macrophage dominated foreign body type reaction,

235  three other histological entities can be readily identified in failed MoM hip replacements, all of

236  which also differ at the macroscopic level. Our results suggest that traditional and often

237  acclaimed "ALVAL-type" responses may be present in two different histological entities, a

238  finding that coincides with the recent consensus statement. These subtypes are naturally very

239  similar in nature: both evince perivascular lymphocyte cuffs, and a high grade of necrosis.

240  Furthermore, both diffuse lymphocyte infiltration, and plasma cells are commonly seen.

241  Moreover, thick-walled PTs are a prominent manifestation in these subtypes. The major

242  difference between these subtypes, based on the results of the current study, is the particle load

243  within histiocytes and extracellular metal particles. One subtype lacked extracellular metal

244  particles, and particle load within histiocytes was low or absent. This subtype also evinced a very

245  high grade of synovial necrosis, and germinal centers were more common in this subtype than in

246  other subtypes. We therefore suggest that this subtype represents the "immunologic type IV

247  neosynovitis" reaction, i.e., true hypersensitivity type reaction. On the other hand, the other,

248  "abrasion induced inflammatory lymphocytic type I neosynovitis", can be regarded as an

249  "ALVAL-type response" associated with wear. In this subtype, the histopathological findings are

250  slightly milder compared with "immunologic type IV neosynovitis". The other groups in our

251  analysis that coincide with the consensus statement were "abrasion-induced foreign body type I

252  neosynovitis" and "abrasion induced necrotic Type I neosynovitis". Further study should validate

253  our preliminary findings in another type of MoM hip replacements.

254

Although the "immunologic type IV neosynovitis" group evinced a high median level of cobalt

ions, it is noteworthy that hips in this group were dominantly HRs. Lehtovirta et al. reported that

circulating metal ion levels correlate poorly with periprosthetic tissue metal content in HRs (22).

THAs most commonly evince taper wear and material loss which is different from debris

originating from the bearing surfaces. Bearing wear may be considered as more fine-grained

resulting in an elevated circulatory level of metal ions but lower tissue metal content. It is

possible that HRs elicit more easily an immunologic neosynovitis type reaction, whereas the

coarser taper wear in THAs is associated with wear-type synovial responses, namely "foreign-

body", "necrotic", and "inflammatory lymphocytic" neosynovitis. It should be, however, noted

that as recent simulator study showed, HRs with edge-loading and microseparation produce

larger and coarser wear particles, possible even larger extracellular particle, compared to more

normal wear conditions (23). Hence "foreign-body", "necrotic", and "inflammatory lymphocytic"

neosynovitis are observed in HRs also.

The current literature on the correlation between wear and ALVAL-reaction is inconsistent.

ALVAL-reaction has been associated with low wear in several studies (6),(24). The authors who

developed the ALVAL-score, which combines three histopathological domains, later showed

that the ALVAL-score does not correlate with bearing wear, and thus contradicts the original

suggestion of the dualistic response of the periprosthetic tissues to wear (25). Grammatopoulos et

al. stated that ALVAL-reaction correlated moderately with increasing wear (24). In support of this

general trend, we observed the highest median levels of metal ions in the immunologic type IV

neosynovitis group, which resemble the "true" ALVAL-reaction. Moreover, Grammatopoulos et

277 al. found that hips with minimal wear and pseudotumour had the most severe ALVAL-reaction.

278 Our results, which suggest that failed MoM hips pose different entities and not just ALVAL and

279 foreign type, also varying in etiology, offer an explanation for the inconsistent findings between

280 the wear and ALVAL response reported in previous studies. Firstly, ALVAL-response or

281 ALVAL-type response may be the result of two different entities as suggested in our study,

282 which may also differ in wear characteristics. Since an ALVAL-type response does not result

283 from a single cascade of response, previous studies may have conflicting results. Secondly, as

284 stated in the consensus statement, different types of neosynovitis are not pathognomic for

285 different etiologies, such as toxic, inflammatory, and immunological ones. "Immunologic type

286 IV neosynovitis" was the most common in our study.  It is unlikely that a pure hypersensitivity

287 response, immunological or allergy, would have solely resulted in such a high number of cases

288 with pure immunologic response. All our patients had been operated on with a recalled high-risk

289 MoM hip replacement that was prone to high wear, and high wear is likely to be present in the

290 majority of our failed hips. We suggest therefore that hypersensitivity and immunologic response

291 to metal wear debris is an important cause of ALVAL-type and other responses. However, in a

292 subset of patients, there is a certain threshold of wear required after which a traditional ALVAL-

293 type reaction starts to develop.

294

295 Coagulative necrosis and perivascular lymphocytes are traditional finding in the ALVAL-type

296 response seen in failed MoM hips. Howie and Vernon-Roberts observed that perivascular

297 lymphocytes are recruited after an intraarticular injection of cobalt chrome particles (26). They

298 suggested that these particles possibly cause a vasculitis-type process or blood vessel occlusion

299 that results in ulceration of the synovial surface and then the recruitment of lymphocytes. This

300    finding is contrary to a study by Witzleb et al. who found no association between the extent of

301    surface necrosis and perivascular lymphocytes (9). Hence, these two finding may develop in

302    isolation and are not pathognomic for each other. In our study, these two responses were seen in

303    three subtypes in variable degrees. Our results also suggest that these three pathological

304    conditions are different, and thus may differ in their etiology. Macroscopically, thick walled PTs

305    were more common in the immunologic and inflammatory neosynovitis groups, whereas PTs

306    were mainly absent in the necrotic neosynovitis group. It is important to realize, however, that

307    two different histopathological phenomena can occur simultaneously. As Mahendra et al. have

308    also suggested, a subgroup of patients may have both ALVAL-type and necrotic reactions that

309    result in implant failure (10).

310

311    A recent study from Ricciardi et al. was very similar to ours (7). The major difference was that

312    they also included non-MoM hips and aimed to investigate the subtypes of histopathological

313    findings related to the corrosion products released from a variety of hip replacements. They

314    defined four subtypes based on the available literature. One of the four subtypes was the

315    macrophage-dominated pattern. The second subtype was "mixed lymphocytic and macrophagic

316    with or without features associated with hypersensitivity/allergy or response to particle toxicity".

317    This subtype is equal to the traditional ALVAL-type reaction that we suggested was dualistic in

318    nature. Ricciardi et al. also suggested the division of this subtype based on the presence of

319    hypersensitivity features. The third subtype described by Ricciardi was predominantly sarcoid-

320    like. Our analysis did not, however, suggest this as a separate entity. This could be due to several

321    factors. Firstly, the majority of cases with sarcoid-like pattern were seen in the non-MoM hips

322    with dual modular taper. We do not suggest therefore that sarcoid-like pattern would constitute a

323     separate entity in MoM hips. Secondly, the segregation methods used in our study rely on the

324     association of histopathological variables. Therefore, if the sarcoid-like granulomas develop in

325     isolation, it is unlikely that this entity would be identified in the segregation analysis.

326

327     We acknowledge some limitations in our study. The most important factor is the possibility that

328     each revised hip may evince several different inflammatory responses. Therefore, there is some

329     overlapping and inconsistencies when the distribution of each histological variable among

330     clusters is interpreted. For example, there are cases with a minor grade of necrosis in the

331     "Cytotoxic" group even though we suggested that the hallmark of this group is synovial necrosis.

332     Cluster analysis does not allow overlapping groups. Hence, cases presenting more than one

333     possible different response are forced to one of the remaining clusters. These cases cannot be

334     identified during the clustering process, and they are merged to other groups. Since cluster

335     analysis readily revealed different groups, however, we do not consider the overlapping of

336     patterns to have been a significant problem.  The prevalence of histological variables may not,

337     however, be representative of true prevalence since they do not develop in isolation from each

338     other. Secondly, the samples obtained from soft tissues perioperatively might not represent the

339     overall response of the synovia. It is not known to what extent one sample of synovia represents

340     the actual type of histopathology in each case. Several samples would minimize this variation,

341     but this approach has practical limitations since tissue preservation is important during revision

342     surgery. Furthermore, there might be considerable variation of cell counts among different

343     sections from the same sample. However, we think that this sampling bias was reduced

344     significantly due to the large group of cases. Finally, we were not able to obtain wear

345     measurements from all our hips. We have published before wear measurement in a subset of

346 ASR patients (27). Median annual wear rate was 9.0 m3. These values are remarkably similar to

347 those reported by Park et al. in a largest cohort of retrieved ASR hips (28). Hence it is unlikely

348 that different wear patterns would have been present in our cohort.

349

350 Our novel findings support the assumption that the microscopic tissue response of failed MoM

351 hips is not just one wide entity with variable inflammatory response and synovial necrosis.

352 Instead, these hips evince four different patterns of histopathological findings. Our study was

353 exploratory in nature, and therefore it requires a follow-up study to investigate how our findings

354 can be reproduced. Further research is also warranted to examine the clinical manifestation of

355 different histopathological patterns in order to enlighten the etiopathogenesis of the adverse soft

356 tissue reactions seen around failed MoM hips and to clarify the association of our findings with

357 those seen in clinical decision making, i.e., blood metal ion levels, hardware related factors, and

358 cross-sectional imaging findings. Additionally, the association between each failure pattern with

359 wear characteristics must be investigated.

360

361 **References**

362 1. Lainiala OS, Reito AP, Nieminen JJ, Eskelinen AP. Declining Revision Burden of Metal-

363      on-Metal Hip Arthroplasties. J Arthroplasty. 2019;

364 2. Hip, Knee & Shoulder Arthroplasty: 2017 Annual Report. 2017;

365 3. National Joint Registry for England and Wales. NJR 14th Annual Report 2017. 2017;

366 4. Willert HG, Buchhorn GH, Fayyazi A, Flury R, Windler M, Koster G, Lohmann CH.

367      Metal-on-metal bearings and hypersensitivity in patients with artificial hip joints. A

368      clinical and histomorphological study. J bone Jt surgeryAmerican Vol. 2005;87:28–36.

369    5.    Davies AP, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic

370          perivascular infiltration in tissues around contemporary metal-on-metal joint

371          replacements. J bone Jt surgeryAmerican Vol. 2005;87:18–27.

372    6.    Campbell P, Ebramzadeh E, Nelson S, Takamura K, Smet K De, Amstutz HC.

373          Histological features of pseudotumor-like tissues from metal-on-metal hips. Clin Orthop

374          Relat Res. 2010;468:2321–7.

375    7.    Ricciardi BF, Nocon AA, Jerabek SA, Wilner G, Kaplowitz E, Goldring SR, Purdue PE,

376          Perino G. Histopathological characterization of corrosion product associated adverse local

377          tissue reaction in hip implants: a study of 285 cases. BMC Clin Pathol. 2016;16:3–016–

378          0025–9. eCollection 2016.

379    8.    Langton DJ, Jameson SS, Joyce TJ, Hallab NJ, Natu S, Nargol A V. Early failure of

380          metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: A

381          consequence of excess wear. J bone Jt surgeryBritish Vol. 2010;92:38–46.

382    9.    Witzleb WC, Hanisch U, Kolar N, Krummenauer F, Guenther KP. Neo-capsule tissue

383          reactions in metal-on-metal hip arthroplasty. Acta Orthop. 2007;78:211–20.

384    10.   Mahendra G, Pandit H, Kliskey K, Murray D, Gill HS, Athanasou N. Necrotic and

385          inflammatory changes in metal-on-metal resurfacing hip arthroplasties. Acta Orthop.

386          2009;80:653–9.

387    11.   Mittal S, Revell M, Barone F, Hardie DL, Matharu GS, Davenport AJ, Martin RA, Grant

388          M, Mosselmans F, Pynsent P, Sumathi VP, Addison O, Revell PA, Buckley CD.

389          Lymphoid aggregates that resemble tertiary lymphoid organs define a specific

390          pathological subset in metal-on-metal hip replacements. PLoS One. 2013;8:e63470.

391    12.   Samelko L, Caicedo MS, Lim S-J, Della-Valle C, Jacobs J, Hallab NJ. Cobalt-alloy

392   implant debris induce HIF-1α hypoxia associated responses: a mechanism for metal-

393   specific orthopedic implant failure. PLoS One. Public Library of Science; 2013;8:e67127.

394 13. Krenn V, Morawietz L, Perino G, Kienapfel H, Ascherl R, Hassenpflug GJ, Thomsen M,

395   Thomas P, Huber M, Kendoff D, Baumhoer D, Krukemeyer MG, Natu S, Boettner F,

396   Zustin J, Kölbel B, Rüther W, Kretzer JP, Tiemann A, Trampuz A, Frommelt L, Tichilow

397   R, Söder S, Müller S, Parvizi J, Illgner U, Gehrke T. Revised histopathological consensus

398   classification of joint implant related pathology. Pathol - Res Pract. 2014;210:779–86.

399 14. Hart AJ, Satchithananda K, Liddle AD, Sabah SA, McRobbie D, Henckel J, Cobb JP,

400   Skinner JA, Mitchell AW. Pseudotumors in association with well-functioning metal-on-

401   metal hip prostheses: a case-control study using three-dimensional computed tomography

402   and magnetic resonance imaging. J bone Jt surgeryAmerican Vol. 2012;94:317–25.

403 15. Reito A, Puolakka T, Elo P, Pajamaki J, Eskelinen A. Outcome of Birmingham hip

404   resurfacing at ten years: role of routine whole blood metal ion measurements in screening

405   for pseudotumours. Int Orthop. 2014;38:2251–7.

406 16. Reito A, Puolakka T, Elo P, Pajamäki J, Eskelinen A. High prevalence of adverse

407   reactions to metal debris in small-headed ASRTM hips. Clin Orthop Relat Res. 2013;471.

408 17. Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol A V. Adverse reactions to metal

409   debris: histopathological features of periprosthetic soft tissue reactions seen in association

410   with failed metal on metal hip arthroplasties. J Clin Pathol. 2012;65:409–18.

411 18. Schreiber JB. Latent Class Analysis: An example for reporting results. Res Soc Adm

412   Pharm. 2017;13:1196–201.

413 19. Beckstead JW. Using Hierarchical Cluster Analysis in Nursing Research. West J Nurs

414   Res. 2002;24:307–19.

415    20.    Gower JC. A General Coefficient of Similarity and Some of Its Properties. Biometrics.

416            International Biometric Society; 1971;27:857.

417    21.    Matthies AK, Racasan R, Bills P, Blunt L, Cro S, Panagiotidou A, Blunn G, Skinner J,

418            Hart AJ. Material loss at the taper junction of retrieved large head metal-on-metal total hip

419            replacements. J Orthop Res. 2013;31:1677–85.

420    22.    Lehtovirta L, Reito A, Parkkinen J, Peräniemi S, Vepsäläinen J, Eskelinen A. Association

421            between periprosthetic tissue metal content, whole blood and synovial fluid metal ion

422            levels and histopathological findings in patients with failed metal-on-metal hip

423            replacement. PLoS One. 2018;13.

424    23.    Kovochich M, Fung ES, Donovan E, Unice KM, Paustenbach DJ, Finley BL.

425            Characterization of wear debris from metal-on-metal hip implants during normal wear

426            versus edge-loading conditions. J Biomed Mater Res Part B Appl Biomater.

427            2018;106:986–96.

428    24.    Grammatopoulos G, Pandit H, Kamali A, Maggiani F, Glyn-Jones S, Gill HS, Murray

429            DW, Athanasou N. The correlation of wear with histological features after failed hip

430            resurfacing arthroplasty. J bone Jt surgeryAmerican Vol. 2013;95:e81.

431    25.    Ebramzadeh E, Campbell P, Tan TL, Nelson SD, Sangiorgio SN. Can wear explain the

432            histological variation around metal-on-metal total hips? Clin Orthop Relat Res.

433            2015;473:487–94.

434    26.    Howie DW, Vernon-Roberts B. The synovial response to intraarticular cobalt-chrome

435            wear particles. Clin Orthop Relat Res. 1988;(232):244–54.

436    27.    Lehtovirta L, Reito A, Parkkinen J, Hothi H, Henckel J, Hart A, Eskelinen A. Analysis of

437            bearing wear, whole blood and synovial fluid metal ion concentrations and

438   histopathological findings in patients with failed ASR hip resurfacings. BMC

439   Musculoskelet Disord. 2017;18.

440 28. Park S-H, Lu Z, Hastings RS, Campbell PA, Ebramzadeh E. Five Hundred Fifty-five

441   Retrieved Metal-on-metal Hip Replacements of a Single Design Show a Wide Range of

442   Wear, Surface Features, and Histopathologic Reactions. Clin Orthop Relat Res.

443   2018;476:261–78.

444

445 **Figure legends**

446 **Figure 1:** Flow chart of case selection and indications for revision. Mechanical reasons indicate

447 fracture, aseptic loosening and dislocation. Extra-articular reasons indicate osteonecrosis (HR

448 only) and heterotopic ossification and rash.

449 **Figure 2**: The upper graph shows the changes in agglomeration coefficient in hierarchical

450 clustering for a defined number of clusters (groups). The middle graph shows the Akaikes

451 information criterion for the defined number of latent classes (groups).

452 **Figure 3**: An example of "abrasion induced necrotic Type I neosynovitis". Synovial lining is

453 destroyed, mild focal necrosis and disruption of synovial surface is seen, histiocytes contain a lot

454 of nanometer scaled metal particles and only few diffuse lymphocytes are present.

455 **Figure 4**: An example of "immunologic type IV neosynovitis". Metal particles and histiocytes

456 are abundant. No extracellular metal particles are seen. Fibrin deposition and ulceration of

457 synovial surface is present as are diffuse and perivascular lymphocyte aggregates.

# PUBLICATION
# IV

**Host-specific factors affect the pathogenesis of adverse reaction to metal debris**

Lehtovirta, L., Reito, A., Lainiala, O., Parkkinen, J., Hothi, H., Henckel, J., Hart, A., & Eskelinen, A.

**BMC Musculoskeletal Disorders**

**Open Access**

# Host-specific factors affect the pathogenesis of adverse reaction to metal debris

Lari Lehtovirta[1,2], Aleksi Reito[2], Olli Lainiala[2], Jyrki Parkkinen[3], Harry Hothi[4], Johann Henckel[4], Alister Hart[4] and Antti Eskelinen[2*]

## Abstract

**Background:** Adverse Reaction to Metal Debris (ARMD) is a major reason for revision surgeries in patients with metal-on-metal (MoM) hip replacements. Most failures are related to excessively wearing implant producing harmful metal debris (extrinsic factor). As ARMD may also occur in patients with low-wearing implants, it has been suggested that there are differences in host-specific intrinsic factors contributing to the development of ARMD. However, there are no studies that have directly assessed whether the development of ARMD is actually affected by these intrinsic factors.

**Methods:** We included all 29 patients (out of 33 patients) with sufficient data who had undergone bilateral revision of ASR MoM hips (58 hips) at our institution. Samples of the inflamed synovia and/or pseudotumour were obtained perioperatively and sent to histopathological analysis. Total wear volumes of the implants were assessed. Patients underwent MARS-MRI imaging of the hips preoperatively. Histological findings, imaging findings and total wear volumes between the hips of each patient were compared.

**Results:** The difference in wear volume between the hips was clinically and statistically significant (median difference 15.35 mm$^3$, range 1 to 39 mm$^3$, IQR 6 to 23 mm$^3$) ($p < 0.001$). The median ratio of total wear volume between the hips was 2.0 (range 1.09 to 10.0, IQR 1.67 to 3.72). In majority of the histological features and in presence of pseudotumour, there were no differences between the left and right hip of each patient ($p > 0.05$ for all comparisons). These features included macrophage sheet thickness, perivascular lymphocyte cuff thickness, presence of plasma cells, presence of diffuse lymphocytic infiltration and presence of germinal centers.

**Conclusions:** Despite the significantly differing amounts of wear (extrinsic factor) seen between the sides, majority of the histological findings were similar in both hips and the presence of pseudotumour was symmetrical in most hips. As a direct consequence, it follows that there must be intrinsic factors which contribute to the symmetry of the findings, ie. the pathogenesis of ARMD, on individual level. This has been hypothesized in the literature but no studies have been conducted to confirm the hypothesis. Further, as the threshold of metal debris needed to develop ARMD appears to be largely variable based on the previous literature, it is likely that there are between-patient differences in these intrinsic factors, ie. the host response to metal debris is individual.

**Keywords:** Metal-on-metal, ASR, ARMD, ALVAL, Pseudotumor, Pseudotumour, Patient susceptibility, Host response

* Correspondence: antti.eskelinen@coxa.fi
Investigation performed at Coxa Hospital for Joint Replacement, Tampere, Finland
[2]Coxa Hospital for Joint Replacement, Tampere, Finland
Full list of author information is available at the end of the article

## Background

Adverse Reaction to Metal Debris (ARMD) continues to be a major reason for revision surgeries in patients with Metal-on-Metal (MoM) hip replacements [1, 2]. ARMD consists of very variable and heterogenous findings and symptoms. Patients may experience strong pain and discomfort or be completely asymptomatic [3]. Radiologically, fluid-filled cystic lesions or solid inflammatory soft-tissue masses termed pseudotumors can be found on some patients, both symptomatic and asymptomatic [4, 5]. Microscopical findings in periprosthetic tissue range from mild macrophage infiltration to deep soft-tissue necrosis with heavy lymphocyte infiltration [6–8]. In summary, there is a high between-subject variability with regard to symptoms, clinical findings and histological presentation of the tissues in patients with ARMD.

Factors that affect the development of ARMD can be divided into extrinsic and intrinsic. The amount of wear debris and physicochemical properties of the particles are examples of extrinsic factors. Intrinsic factors, such as individual differences in innate and adaptive immune responses to metal wear debris, can be collectively referred to as host response [9]. Several retrieval studies have investigated extrinsic factors, most importantly implant wear, and their association to ARMD. Many studies have shown that implant wear is a risk factor for the development of ARMD [10–12]. However, adverse reactions have also been observed in patients with low wearing hip implants in several studies [8, 13–15]. In their systematic review, Campbell et al. concluded that no clear dose-response relationship between wear and ARMD could be established due to the heterogeneity of the findings in the included studies [16]. Studies that have investigated association between the histopathological features of ARMD and wear or indirect markers of wear, such as synovial fluid or whole blood metal ion concentrations, have also yielded inconsistent results [6, 8, 14, 17–23]. The lack of a clear association between extrinsic factors and the development of ARMD could be due to a remarkable role of intrinsic factors affecting the pathogenesis. In fact, the contribution of host-specific factors and presence of patient susceptibility has been suggested in numerous previous studies based on the between-subject discrepancy in the amount of wear debris needed to result in ARMD and implant failure [8, 13, 14, 24–28]. Further, it has been suggested that women are more susceptible than men, possibly due to previous exposure to metals from jewelry [14, 15, 29]. However, to the best of our knowledge, there are no studies that would have actually investigated whether intrinsic factors affect the pathogenesis of ARMD in patients with MoM hips.

In the present study, we aimed to indirectly investigate whether host-specific intrinsic factors affecting the pathogenesis of ARMD exist in a cohort of patients with bilateral ASR hips, both of which were revised for ARMD. Host response was investigated by comparing both histological findings and the amount of bearing surface wear volume (extrinsic factor) between each patient's left and right hips. Each hip served as a control for the other. If the tissue response between the hips was similar (low within-subject variability) despite differing amount of wear debris between the sides (difference in an extrinsic factor), it would indicate the presence of intrinsic factors contributing to the similarity of the tissue response (Additional file 1). We had three hypotheses: 1) there is significant congruence in histological findings between the hips of each patient (low within-subject variability) despite differing amount of wear between the hips, indicating the contribution of intrinsic factors in the pathogenesis, 2) histological findings characteristic of the innate immune response or direct cytotoxic effects of metal debris (macrophages, granulomas and necrosis) would differ between the sides in response to wear debris and 3) components of the individual adaptive immune response (lymphocytes, germinal centers and plasma cells) would be congruent between the sides as a result of contribution of intrinsic factors.

## Methods

### Study design

One thousand thirty-six Articular Surface Replacement (ASR) MoM hip replacements (Depuy Orthopaedics, Warsaw, IN, USA) were performed in 887 patients at our institution between March 2004 and December 2009. By the end of September 2016, 316 patients had been revised. Of these, 33 patients have undergone bilateral revision. Four of these patients were excluded due to missing tissue samples thus leading to 29 patients being included in our study (58 hips). Flow chart of the patient selection is available as a supplement (Additional file 2). All patients had the same head-cup-combination on both sides: five patients had bilateral ASR hip resurfacing and 24 patients had ASR XL stemmed total hip replacements bilaterally. Simultaneous bilateral hip revision was performed for two patients, and the remaining 27 patients' bilateral revision surgeries were performed sequentially. Revision operations have been described in detail in our previous publication [30]. Patient demographics and indications for revision surgery are presented in Table 1. Surgery was performed by or under the direct supervision of 10 senior orthopedic surgeons. All patients gave written informed consent to participate in this study that was approved by the ethical committee of Pirkanmaa Hospital District (R11006).

### Follow-up

After the recall of DePuy ASR hip arthroplasties and the Medicines and Healthcare products Regulatory Agency

**Table 1** Reasons for revision surgery

| Reasons for revision surgery | |
| --- | --- |
| Progressively elevating whole blood metal ion levels | 22 hips (38%) |
| Symptomatic hip and elevated whole blood metal ion levels | 14 hips (24%) |
| Symptomatic hip, not elevated whole blood metal ion levels | 5 hips (9%) |
| Pseudotumor and elevated whole blood metal ion levels | 14 hips (24%) |
| Aseptic cup loosening | 3 hips (5%) |
| Total | 58 hips (100%) |

(MHRA) medical device alert regarding MoM hip arthroplasties, a systematic screening programme was launched at our institution [31, 32]. All patients with MoM hip arthroplasty were included in the programme. Patients were given Oxford Hip Score questionnaire, examined physically (including the Harris Hip Score) and whole blood chromium and cobalt ion levels were measured [33, 34]. Hip and pelvic radiographs were taken before each visit. In addition, all patients were referred for Metal Artifact Reduction Sequence MRI (MARS-MRI), unless there were contraindications, in which case patients were referred for ultrasound imaging of the hips. Findings were classified using a previously published pseudotumour classification [4]. For the purposes of the study, pseudotumours were considered as fluid-filled or solid soft-tissue masses adjacent to the articulation (classes 1, 2A, 2B or 3).

### Indications for revision surgery

Revision surgery was considered if 1) a clear pseudotumour (class 2A,2B or 3) [4] was observed on cross-sectional imaging regardless of symptoms or whole blood (WB) metal ion levels; or 2) the patient had elevated WB metal ion levels and hip symptoms despite normal findings in cross-sectional imaging; or 3) the patient had a continuously symptomatic hip or progressive symptoms regardless of imaging findings or metal ion levels; or 4) the patient had progressively increasing blood metal ion levels, even without symptoms or findings in cross-sectional imaging. Symptoms included hip pain, discomfort, sense of instability, and/or impaired function of the hip and sounds from the hip (clacking, squeaking). WB metal ion levels were regarded as being elevated if either chromium or cobalt exceeded 5 ppb [35].

### Bearing wear analysis

The volume of material loss from the cup and head bearing surfaces was measured using a Zeiss Prismo (Carl Zeiss Ltd., Rugby, UK) coordinate measuring machine (CMM). A total of 400 polar scan lines on each surface were defined and up to 30,000 data points captured using a 2 mm ruby stylus; protocols for this method have been previously published [36]. An iterative least square fitting method was used to analyze the raw

data captured by the CMM and the unworn geometry of the bearing surface was used to map regions of material loss from which the total volumetric loss was calculated for each component. Total wear volume was calculated by combining head and cup wear volumes for each patient.

### Histopathological analysis of the periprosthetic tissue

During every hip revision, samples of the inflamed synovia or pseudotumor capsule were obtained. For histopathological analysis, each tissue sample was formalin fixed and embedded in paraffin. Several 10 μm microtome sections were made and stained with standard hematoxylin and eosin staining. The sections were examined histologically under transmitted light with a Nikon Eclipse 50i microscope (Nikon Corporation, Shinagawa, Tokyo, Japan). The sections were graded by a senior musculoskeletal pathologist (JP) using scoring principles adopted from the study by Natu et al. [7]. The pathologist was blinded from clinical patient characteristics.

The Natu grading consisted of following parameters: 1) macrophage sheet thickness, 2) lymphocyte cuff thickness, 3) degree of necrosis, 4) presence of plasma cells, 5) presence of diffuse lymphocytic infiltrate, 6) presence of germinal centers, and 7) presence of granulomas. Thickness of histiocyte sheets was calculated using a graticule and graded 0–3 (absent, < 1 mm, 1–2 mm, > 2 mm). Lymphocyte cuff thickness was also calculated using a graticule. An average of five measurements was taken and graded as 0–3 (absent, 0.25 mm, 0.25–0.75 mm, > 0.75 mm). The extent of overall tissue necrosis in a sample was graded based on the surface necrosis typing according to Davies et al. [37]. Type 1 surface contains intact synovial epithelium. Type 2 surface shows loss of synovial epithelial cells without fibrin deposition. In type 3 surface there is fibrin deposition and in type 4 surface there is extensive necrosis and loss of architecture. The extent of type 4 surface necrosis was used to grade the overall tissue necrosis in a given sample, as described by Natu et al. [7]. In grade 4 necrosis, more than 75% of the tissue sample showed type 4 surface necrosis. In grade 3 necrosis, between 25 and 75% showed type 4 surface necrosis. In grade 2 necrosis either less than 25% of the tissue showed type 4 surface necrosis or the tissue showed type 3 surface. In grade 1 necrosis, the sample consisted of type 2 surface.

### Statistical analysis

Statistical analyses were performed using SPPS software (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). Medians, ranges and interquartile ranges were calculated for total wear volume in both hips (skewed distribution). The statistical significance of the difference in wear volume between the higher and lower wearing side

was tested using Wilcoxon signed ranks test (related samples). Mann-Whitney U-test was used to test the difference in wear volume distribution between the hips in patients with symmetric versus asymmetric histological and imaging findings (independent samples). The differences in histological findings between left and right hips were compared and number of patients with identical findings, patients with a difference of one point, difference of two points between the sides etc. calculated. The statistical significance of the difference in histological findings between the sides was tested with marginal homogeneity test except the difference in presence of germinal centers which did not fill the test requirements and McNemar test was used instead [38]. Whether presence of MRI-confirmed pseudotumour was similar between left and right sides was tested using McNemar test (related samples).

## Results
Thirteen of the 29 patients included in the study were females (45%). Mean age of the patients was 61.7 years (SD 8.3 years) at the time of the first revision operation and 63.1 years (SD 8.5 years) at the time of the second revision operation, respectively. On average, the first hip was revised 4.5 years (SD 1.29 years) and the second hip 5.8 years (SD 1.8 years) after the primary operation.

Component wear was available bilaterally for 17 (59% of all) patients. Total wear volume in either hip ranged from $3\,mm^3$ to $94\,mm^3$ (median $13\,mm^3$, IQR 10 to $32\,mm^3$). The median difference in wear volume between higher and lower wearing side was $15.35\,mm^3$ (range 1 to $39\,mm^3$, IQR 6 to $23\,mm^3$) ($p < 0,001$). This difference is illustrated in Fig. 1. The median ratio of total wear volume between the hips was 2.0 (range 1.09 to 10.0, IQR 1.67 to 3.72). In 9 of the 17 (53%) patients with wear data available, the ratio of wear was 2.0 or greater, ie. there was at least two-fold difference in the wear volume between the hips.

The variability of histological findings was high (Table 2). Most hips evinced mild-to moderate macrophage and lymphocyte infiltration, while in some patients there was heavy infiltration of either macrophages or lymphocytes but not both simultaneously. The degree of necrosis was approximately evenly distributed in all five grades. Majority of patients evinced no plasma cells, diffuse lymphocytic infiltration, germinal centers or granulomas.

The congruence of histological findings between the left and the right hips is presented in Table 3. In majority of the histological features and also in majority of the patients, there were no differences between the hips ($p > 0.05$ for all comparisons). These features included macrophage sheet thickness, perivascular lymphocyte cuff thickness, presence of plasma cells, presence of diffuse lymphocytic infiltration and presence of germinal

centers. In lymphocyte cuff thickness the difference between the sides was at most 1 point. In macrophage sheet thickness the findings were similar in 18 patients, differed by 1 point in 9 patients and differed by 2 points in 2 patients, respectively. The only histological findings that statistically significantly differed between the hips were grade of necrosis ($p < 0.01$) and presence of granulomas ($p = 0.025$). In the grade of necrosis there was a wide distribution in the difference between the sides. In those patients with granuloma present on one side only, the granuloma was always on the higher-wearing side. When comparing all hips, those hips with a granuloma ($n = 5$) had a median total wear volume of $35\,mm^3$ (range 15.0 to 111.0) and those hips with no granuloma ($n = 39$) had a median total wear volume of $15\,mm^3$ (range 3.0 to 94.0) ($p = 0.059$ for comparison). In the grade of necrosis, the higher grade was not always on the higher wearing side. In any of the histological findings, the symmetry or asymmetry of findings between left and right sides was not associated with a difference in the distribution of wear volume between the sides (Table 4). All patients had at least two histological variables with similar findings on both hips. Majority of the patients (75.9%) had four or more histological variables with similar findings on both sides (Table 5). There were no differences in the similarity or dissimilarity of histological findings between left and right hips in males versus females (Table 6).

Bilateral MRI classification for the presence of pseudotumours was available for 25 patients (86% of all patients). 18 patients (72% of the classified) had either bilateral pseudotumours or no pseudotumours at all on either side, ie. the hips were symmetrical in regard to pseudotumour. There was no statistically significant difference in the presence of pseudotumour between the sides ($p = 0.13$). Of those 18 patients, 7 had pseudotumour on both sides (of which two were identical by exact classification) and 11 had no pseudotumour on either side. Patients with asymmetrical pseudotumour finding between the sides evinced similar distribution of total wear volume between the sides as those patients with symmetrical pseudotumour findings (Table 7). In addition, there were no differences in the total wear volumes of the hips in patients with pseudotumour on both sides (median $20.0\,mm^3$, range 9.0 to 111.0) versus no pseudotumour on either side (median $16.30\,mm^3$, range 3.0 to 51.0) ($p = 0.28$ for comparison).

## Discussion
In the present study, we found that there were notable differences in the histological findings between patients revised for ARMD, ie. the between-subject variability was high. Heterogeneity has been characteristic for the results of ARMD research [16]. Most importantly,
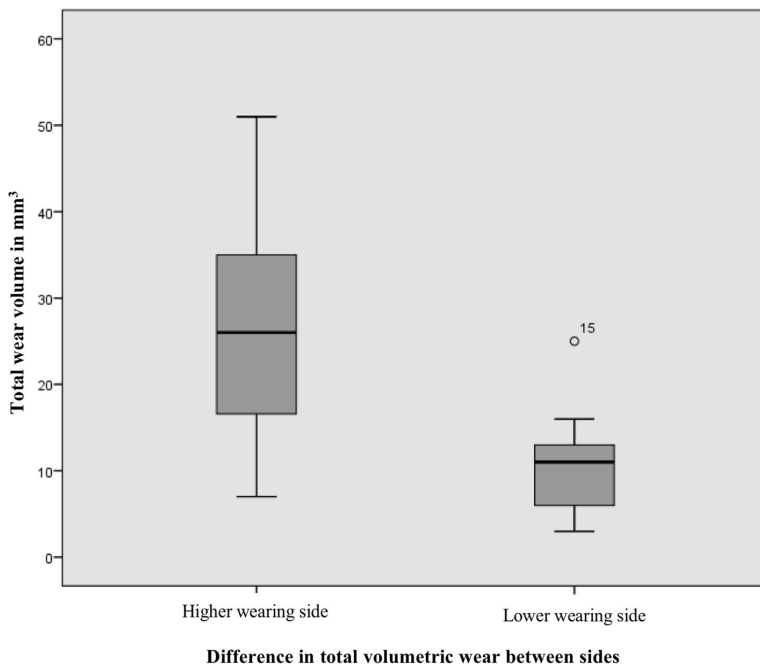
**Fig. 1** The difference in total wear volume between higher and lower wearing sides

however, we found no statistically or clinically significant differences in most of the histological and imaging findings between left and right hips of the same patient, meaning that the within-subject variability in histological and imaging findings was low. Further, majority of the patients had similar findings on both hips in several key histological variables. This was despite the fact that there was a clinically and statistically significant difference in the amount of wear volume between the sides, ie. there was a difference in the extrinsic factor between the sides. There are no clearly defined boundaries for abnormal versus normal wear, but volumetric wear rates exceeding $1 \, mm^3$/year are generally considered abnormal [39]. As the median difference of $15.4 \, mm^3$ in wear volume between the sides measured in our study translates into remarkably abnormal yearly volumetric wear rate needed to generate that difference, we thus feel safe to consider the difference in median wear volume between the sides clinically significant.

The contribution of host-specific factors in the pathogenesis of ARMD has been suggested in numerous previous studies, likely observed as patient susceptibility of different levels [8, 13, 14, 24–28]. However, to the best of our knowledge there are no previous studies that would have actually assessed the role of intrinsic factors in the pathogenesis. On the contrary, there are many

studies that have investigated implant wear and the development of ARMD, however, results of these studies are very discrepant. High wear or high blood metal ion levels resulting from high wear are associated with the development of ARMD [10, 40]. However, adverse reactions have been noted in patients with both high and low wearing hip implants [8, 11, 13, 25, 41]. In a systematic review by Campbell et al. no clear dose-response relationship between wear and ARMD could be established [16]. We observed symmetry of histological findings between left and right hips despite differing amounts of wear. In addition, the distribution of wear volume between the sides was similar in patients with symmetrical versus asymmetrical histological and imaging findings. Further, patients with bilateral pseudotumours had similar amounts of wear volumes in their hips as did patients with no pseudotumour on either side. Our finding suggests that there are intrinsic factors that markedly contribute to the pathogenesis of ARMD, dictating the type of tissue response and development of pseudotumours, in addition to extrinsic factors such as volume of the metal wear debris. Further, it is likely that there are differences in these intrinsic factors between patients as some develop aggressive tissue responses despite low-wearing implant while some tolerate large amounts of wear. Various terms have been used to

**Table 2** Between-subject differences in histological findings

| Histological finding | Right hips | Left hips |
|---|---|---|
| Macrophage sheet thickness | | |
| 0 (absent) | 1 (3.4%) | 1 (3.4%) |
| 1 (< 1 mm) | 19 (65.5%) | 24 (82.8%) |
| 2 (1–2 mm) | 7 (24.1%) | 4 (13.8%) |
| 3 (> 2 mm) | 2 (6.9%) | 0 (0.0%) |
| Lymphocyte cuff thickness | | |
| 0 (absent) | 13 (44.8%) | 13 (44.8%) |
| 1 (0.25 mm) | 11 (37.9%) | 11 (37.9%) |
| 2 (0.25–0.75 mm) | 5 (17.2%) | 4 (13.8%) |
| 3 (> 0.75 mm) | 0 (0.0%) | 1 (3.4%) |
| Degree of necrosis | | |
| 0 | 6 (20.7%) | 0 (0.0%) |
| 1 | 4 (13.8%) | 3 (10.3%) |
| 2 | 8 (27.6%) | 7 (24.1%) |
| 3 | 7 (24.1%) | 5 (17.2%) |
| 4 | 4 (13.8%) | 14 (48.3%) |
| Presence of plasma cells | | |
| No | 23 (79.3%) | 22 (75.9%) |
| Yes | 6 (20.7%) | 7 (24.1%) |
| Presence of diffuse lymphocytic infiltration | | |
| No | 22 (75.9%) | 20 (69.0%) |
| Yes | 7 (24.1%) | 9 (31.0%) |
| Presence of germinal centers | | |
| No | 27 (93.1%) | 29 (100%) |
| Yes | 2 (6.9%) | 0 (0.0%) |
| Presence of granulomas | | |
| No | 23 (79.3%) | 28 (96.6%) |
| Yes | 6 (20.7%) | 1 (3.4%) |

describe this phenomenon, for example patient susceptibility [13]. Clinicians should bear in mind that some patients with low wearing implants (low blood metal ion levels) can still be at risk for ARMD due to higher than average patient susceptibility.

A cohort of patients with bilateral MoM hips forms an excellent research frame to investigate and compare the role of intrinsic and extrinsic factors in the pathogenesis. We are aware of only three previous studies that compare characteristics of ARMD between the sides in patients with bilateral MoM hip replacements. Madanat et al. compared MRI findings between left and right hips in patients with bilateral MoM hip replacements [42]. They found that the soft tissue reaction observed in MRI was symmetrical between the sides in most patients, both in sequentially and simultaneously implanted hips. In support of their findings, we report similar symmetry for the presence of MRI-confirmed pseudotumour between the sides. Another study by Pandit et al. consisted of four revised patients with bilateral MoM hips [43]. All patients had developed a necrotic pseudotumor in both hips. In histopathological analysis, both hips of each patient had similar findings (necrosis, macrophages, lymphocytes). However, no wear data was included in the study and the histology was descriptive, not semi-quantitatively scored. A recent study by Uchihara et al. included patients with both uni- and bilateral MoM hips that had been revised for ARMD [44]. They compared histological findings between left and right hips in the bilateral patients as well as histological findings between unilateral and bilateral patients. In addition, time-to-failure was compared between these two groups. The histological findings (necrosis, macrophages, lymphocytes) between left and right hips of the bilateral patients were found to be symmetrical in majority of the cases, similar to the findings of the present study. However, we observed that there were differences in the grade of necrosis between the sides while Uchihara et al. did not semiquantitatively grade the necrosis. Further, there were no differences in the histological findings or time-to-failure between uni- and

**Table 3** Congruence in histological grading between left and right hips (within-subject)

| | Difference in histological grading between left and right sides | | | | | |
|---|---|---|---|---|---|---|
| | No difference | 1 p | 2p | 3p | 4p | Scale |
| Macrophage sheet thickness | 18 (62%) | 9 (31%) | 2 (7%) | – | | 0–3 p |
| Lymphocyte cuff thickness | 14 (48%) | 15 (52%) | – | – | | 0–3 p |
| Degree of necrosis* | 6 (21%) | 10 (34%) | 9 (31%) | 1 (3%) | 3 (10%) | 0–4 p |
| Presence of plasma cells | 26 (90%) | 3 (10%) | | | | Yes/no |
| Presence of diffuse lymphocytic infiltration | 19 (66%) | 10 (34%) | | | | Yes/no |
| Presence of germinal centers | 27 (93%) | 2 (7%) | | | | Yes/no |
| Presence of granulomas* | 24 (83%) | 5 (17%) | | | | Yes/no |

Percentages represent proportion of all patients. In variables marked with * there was a statistically significant ($p < 0.05$) difference between the sides (see results)

**Table 4** Median differences in total wear volumes between the sides (mm³)

| Histology between sides | Symmetrical | Asymmetrical | *P*-value |
|---|---|---|---|
| Macrophages | 9.0 | 18.7 | 0.40 |
| Lymphocytes | 18.0 | 12.7 | 0.89 |
| Necrosis | 32.0 | 12.7 | 0.35 |
| Plasma cells | 16.0 | 5.3 | 0.24 |
| Diffuse lymphocytes | 16.0 | 6.3 | 0.48 |
| Germinal centers | 15.7 | 23.0 | 0.71 |
| Granulomas | 15.35 | 16.1 | 0.70 |

Median differences in total wear volumes between the sides in patients with symmetrical histological findings versus patients with asymmetrical histological findings. Only patients with complete wear data are included (*n* = 17)

bilateral patients in their study. Uchihara et al. concluded that the implantation of a MoM hip does not appear to lead to sensitization to metal debris that would in turn lead to poor clinical performance or different tissue response in the second MoM hip. However, they did not discuss the significance of their findings in the context of intrinsic factors contributing to the similarity of the tissue response between the hips in bilateral patients. Further, their sample size was rather small (10 patients) and no wear data of the MoM hips was presented in the study. These three previous studies conducted on bilateral MoM patients are in agreement with our findings and support the hypothesis of an individual host response dictated by intrinsic factors as a significant contributor in the development of soft tissue reactions leading to failure of the hip.

The pathogenesis of ARMD is poorly understood, but at least three different mechanisms of failure have been suggested: 1) type IV hypersensitivity response to metal wear debris with adaptive immunity involvement, 2) foreign-body response to metal wear particles reflecting innate immunity and 3) direct cytotoxic effect of metal

**Table 5** The degree of similarity between the hips measured by the number of histological variables with similar findings on both sides in each patient

| Histological variables with symmetric findings on both sides | Number of patients | Percentage of patients |
|---|---|---|
| 0 | 0 | 0% |
| 1 | 0 | 0% |
| 2 | 2 | 6.9% |
| 3 | 5 | 17.2% |
| 4 | 6 | 20.7% |
| 5 | 6 | 20.7% |
| 6 | 9 | 31.0% |
| 7 | 1 | 3.4% |
| | Total 29 | Total 100% |

ions [6, 8, 45]. To what degree the tissue response depends on the amount of wear and to what degree on the host-specific intrinsic factors is not well understood. We hypothesized that components of the innate response (macrophages, granulomas, necrosis) are more closely related to extrinsic factors and components of the adaptive response (lymphocytes, germinal centers and plasma cells) to intrinsic factors such as genetic predisposition to metal hypersensitivity. We found that the grade of tissue necrosis and presence of granulomas differed between the sides in most patients. Granulomas were always present on the higher wearing side. Further, when analyzing all hips as a group, we found that there was a trend for higher total wear volume in hips with a granuloma compared to those hips with no granuloma. However, this difference did not quite reach statistical significance. Granulomas are considered to form as a response to high numbers of metal particles in tissues [46]. Our results support this idea. Still, in the present study granulomas were not present in the majority of the hips. We suggest that there is a certain threshold for tissue metal content needed for granulomas to develop as a response. Whether this threshold is dependent on intrinsic factors, particle size, non-particulate metal debris or particle type is not understood and requires further research. The metal ions released from implants are known to cause dose-dependent cytotoxicity in-vitro [47]. Also, we and others have previously shown that implant wear correlates with necrosis of the periprosthetic tissues [6, 48]. Thus, it seems likely that extrinsic factors, mainly implant wear, are more important in the development of tissue necrosis and granulomas than intrinsic factors, ie. patient susceptibility. However, opposite to our hypothesis, the grade of macrophage sheet thickness did not differ between the sides. This would suggest that the macrophage response (innate) is mostly determined by host-specific factors instead of extrinsic factors such as volume of the wear debris. However, there are limitations in our methodology. We did not directly measure the number of macrophages, instead, we measured the thickness of the macrophage sheets. It is possible that the infiltration penetrates deep in the tissue but is not dense. We observed that there were no statistically significant differences in the amounts of lymphocytes and presence of plasma cells and germinal centers between the hips, despite markedly different wear volumes in most of these patients. These parameters belong to the adaptive immune system which is considered host-specific. Thus, it makes sense that they are expressed symmetrically. In some studies, it has been found that low wear is associated to adaptive lymphocytic response and high wear to innate, macrophage dominated foreign-body response [6, 8, 21]. These associations have been weak, however. In addition, disagreeing findings have been

**Table 6** Comparison of similar versus not similar histological findings between the sides in males and females

| Histological variable | Symmetric findings on both hips | Males | Females | P-value for the difference between males and females |
|---|---|---|---|---|
| Macrophage sheet thickness | Yes | 11 (69%) | 7 (54%) | 0.46 |
| | No | 5 (31%) | 6 (46%) | |
| Lymphocytic cuff thickness | Yes | 9 (56%) | 5 (38%) | 0.46 |
| | No | 7 (44%) | 8 (62%) | |
| Degree of necrosis | Yes | 4 (25%) | 2 (15%) | 0.66 |
| | No | 12 (75%) | 11 (85%) | |
| Presence of plasma cells | Yes | 14 (88%) | 12 (92%) | 0.58 |
| | No | 2 (12%) | 1 (8%) | |
| Presence of diffuse lymph. | Yes | 12 (75%) | 7 (54%) | 0.27 |
| | No | 4 (25%) | 6 (46%) | |
| Presence of germinal centers | Yes | 15 (94%) | 12 (92%) | 1.0 |
| | No | 1 (6%) | 1 (8%) | |
| Presence of granulomas | Yes | 15 (94%) | 9 (69%) | 0.14 |
| | No | 1 (6%) | 4 (31%) | |

published [19, 22]. We suggest that host-specificity of the intrinsic factors leads to differences in the tissue response between individuals no matter what the wear. This likely contributes to the poor association between the amount of wear and type of inflammatory tissue response in previous literature.

Our study is not without limitations. First, the sample size in our study is rather small. However, it is clearly the largest in any published study dealing with this issue so far. Second, not all hips were analyzed for bearing wear volume. However, it must be noted that large patient cohorts with clinical information, laboratory and imaging findings, tissue samples and also retrieval analyses available, are not easily available anywhere globally. Further, our patient cohort is free of selection bias as all patients have been primarily operated and followed-up thereafter at our institution with no referrals from other centers. Thirdly, we were not able to analyze the volume of the material loss from the trunnion in those patients with ASR XL hip implants. However, the volume of the material loss from the trunnion is known to be less than that from the bearing couple [49]. Fourth, we used surrogate markers (semiquantitative histology) to indirectly investigate the presence of intrinsic factors contributing to the response. Measuring variability in signaling pathways

**Table 7** Pseudotumour finding and wear volumes between the sides

| Pseudotumour | Symmetrical | Asymmetrical | P-value |
|---|---|---|---|
| Median difference in total wear volume between the sides (mm$^3$) | 12.7 | 13.5 | 0.79 |

The distribution of wear volume between left and right sides is similar in patients with symmetrical and asymmetrical pseudotumour findings between the sides. Only patients with complete wear data are included (n = 17)

provide more direct evidence, but was out of the scope of the current study. Besides, histological methods are well-documented and there is vast amount of literature regarding ARMD histology. However, it is not yet well understood which signaling pathways are important in the development of ARMD and thus a comprehensive study of such would not be realistic. Our study offers novel insight into the role of intrinsic versus extrinsic factors in the pathogenesis of ARMD and is the largest bilateral patient cohort published on the subject. Further, our study is the first one to include wear data.

## Conclusion

In conclusion, intrinsic host-specific factors most likely contribute to the development of ARMD in addition to extrinsic factors such as implant wear debris. Further, it is likely that there are differences in these host-specific factors between patients, manifesting as susceptibility to metal debris of variable degree. Clinicians should bear in mind that patients may have different responses to the same amount of wear debris, usually measured as blood metal ion levels. Some patients may tolerate high amounts of metal debris and some patients may develop even severe adverse tissue responses in the presence of a low-wearing hip implant. Also, bilateral MoM patients with failure on one side will likely develop a similar tissue response on the other side as well. This should be accounted for in the follow-up of patients with bilateral MoM hip replacements. In future studies, it is important to search for possible biomarkers that would predict the severity and type of the intrinsic response, in other words, patient susceptibility. Further, it is important to understand the true nature of ARMD in order to be able to design safer bearing couples in the future.

## Additional files

**Additional file 1:** Between-patient and within-patient variability explained in detail. (DOCX 65 kb)

**Additional file 2:** Flow chart of the patient selection for the study. (TIF 57 kb)

### Authors' contributions
LL formed and analyzed the data and wrote the initial draft of the manuscript. AR, OL, JP, HH, AH, JH and AE assisted in interpretation of the data and editing of the manuscript. All authors read and approved the final manuscript.

### Ethics approval and consent to participate
All patients gave written informed consent and the study was approved by the ethical committee of Pirkanmaa Hospital District (R11006).

### Consent for publication
Not applicable.

### Competing interests
Authors LL, AR, HH and JH have no competing interests related to the study. Authors OL and JP have received lecture fees from DePuy Synthes. Author AH has a research contract with DePuy Synthes. Author AE has received lecture fees from Zimmer Biomet and institutional research funding from DePuy Synthes and Zimmer Biomet.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details
$^{1}$Faculty of Medicine and Health Technology, University of Tampere, Tampere, Finland. $^{2}$Coxa Hospital for Joint Replacement, Tampere, Finland. $^{3}$Fimlab Laboratories Oy, Tampere, Finland. $^{4}$University College London, London, UK.

## References
1. Bosker BH, Ettema HB, Boomsma MF, Kollen BJ, Maas M, Verheyen C. High incidence of pseudotumour formation after large-diameter metal-on-metal total hip replacement a prospective cohort study. J Bone Jt Surgery, Br. 2012;94:755–61.
2. Reito A, Lainiala O, Elo P, Eskelinen A. Prevalence of failure due to adverse reaction to metal debris in modern, medium and large diameter metal-on-metal hip replacements--the effect of novel screening methods: systematic review and Metaregression analysis. PLoS One. 2016;11:e0147872.
3. Browne JA, Bechtold CD, Berry DJ, Hanssen AD, Lewallen DG. Failed metal-on-metal hip arthroplasties: a Spectrum of clinical presentations and operative findings. Clin Orthop Relat Res. 2010;468:2313–20.
4. Hart AJ, Satchithananda K, Liddle AD, Sabah SA, Mcrobbie D, Henckel J, et al. Pseudotumors in Association with Well-Functioning Metal-on-Metal Hip Prostheses. Jbjs-a. 2012;94(4):317–25.
5. Williams DH, Greidanus NV, Masri BA, Duncan CP, Garbuz DS. Prevalence of Pseudotumor in asymptomatic patients after metal-on-metal hip arthroplasty. J Bone Jt Surgery-American. 2011;93:2164–71.
6. Grammatopoulos G, Pandit H, Orth F, Kamali A, Maggiani F, Athanasou N, et al. The correlation of Wear with histological features after failed hip resurfacing arthroplasty. J Bone Jt Surg - Ser A. 2013;81:1–10.
7. Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol AV. Adverse reactions to metal debris: histopathological features of periprosthetic soft tissue reactions seen in association with failed metal on metal hip arthroplasties. J Clin Pathol. 2012;65:409–18.
8. Campbell P, Ebramzadeh E, Nelson S, Takamura K, De Smet K, Amstutz H. Histological features of Pseudotumor-like tissues from metal-on-metal hips. Clin Orthop Relat Res. 2010;468:2321–7.
9. Chaplin DD. Overview of the immune response. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S3–23.
10. Langton DJ, Jameson SS, Joyce TJ, Hallab NJ, Natu S, Nargol AVF. Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: a consequence of excess wear. J Bone Jt Surgery, Br. 2010;92:38–46.
11. Langton DJ, Joyce TJ, Jameson SS, Lord J, Van Orsouw M, Holland JP, et al. Adverse reaction to metal debris following hip resurfacing the influence of component type, orientation and volumetric wear. Bone Joint J. 2011;93(2):164–71.
12. Glyn-Jones S, Pandit H, Kwon YM, Doll H, Gill HS, Murray DW. Risk factors for inflammatory pseudotumour formation following hip resurfacing. J Bone Jt Surg. 2009;91(12):1566–74.
13. Matthies AK, Skinner JA, Osmani H, Henckel J, Hart AJ. Pseudotumors are common in well-positioned low-wearing metal-on-metal hips. Clin Orthop Relat Res. 2012;470:1895–906.
14. Ebramzadeh E, Campbell PA, Takamura KM, Lu Z, Sangiorgio SN, Kalma JJ, et al. Failure modes of 433 metal-on-metal hip implants: how, why, and wear. Orthop Clin North Am. 2011;42:241–50.
15. Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gibbons CL, et al. Pseudotumours associated with metal-on-metal hip resurfacings. J bone Jt surgeryBritish Vol. 2008;90:847–51.
16. Campbell PA, Kung MS, Hsu AR, Jacobs JJ. Do retrieval analysis and blood metal measurements contribute to our understanding of adverse local tissue reactions? Clin Orthop Relat Res. 2014;472:3718–27.
17. Reito A, Parkkinen J, Puolakka T, Pajamäki J, Eskelinen A. Diagnostic utility of joint fluid metal ion measurement for histopathological findings in metal-on-metal hip replacements. BMC Musculoskelet Disord. 2015;16:1.
18. Langton D, Sidaginamale R, Lord J, Joyce T, Natu S, Nargol A. Metal debris release from taper junctions appears to have a greater clinical impact than debris released from metal on metal bearing surfaces. Bone J J. 2013;95 SUPP 1:1.
19. Lohmann CH, Meyer H, Nuechtern JV, Singh G, Junk-Jantsch S, Schmotzer H, et al. Periprosthetic tissue metal content but not serum metal content predicts the type of tissue response in failed small-diameter metal-on-metal total hip arthroplasties. J Bone Jt Surg. 2013;95:1561–8.
20. Nawabi DH, Nassif NA, Do HT, Stoner K, Elpers M, Su EP, et al. What causes unexplained pain in patients with metal-on metal hip devices? A retrieval, histologic, and imaging analysis. Clin Orthop Relat Res. 2014; 472:543–54.
21. Paukkeri E-L, Korhonen R, Hämäläinen M, Pesu M, Eskelinen A, Moilanen T, et al. The inflammatory phenotype in failed metal-on-metal hip arthroplasty correlates with blood metal concentrations. PLoS One. 2016;11:e0155121.
22. Liow MHL, Urish KL, Preffer FI, Nielson GP, Kwon Y-M. Metal ion levels are not correlated with histopathology of adverse local tissue reactions in taper corrosion of total hip arthroplasty. J Arthroplast. 2016;31:1797–802.
23. Langton DJ, Jameson SS, Joyce TJ, Gandhi JN, Sidaginamale R, Mereddy P, et al. Accelerating failure rate of the ASR total hip replacement. J bone Jt surgeryBritish Vol. 2011;93:1011–6.
24. Donell ST, Darrah C, Nolan JF, Wimhurst J, Toms A, Barker THW, et al. Early failure of the ultima metal-on-metal total hip replacement in the presence of normal plain radiographs. J. Bone Joint Surg. 2010;92:1501–8.

Lehtovirta *et al. BMC Musculoskeletal Disorders* (2019) 20:195

Page 10 of 10

25. Ebramzadeh E, Campbell P, Tan TL, Nelson SD, Sangiorgio SN. Can Wear explain the histological variation around metal-on-metal Total hips? Clin Orthop Relat Res. 2014;473:487–94.

26. Hart AJ, Matthies A, Henckel J, Ilo K, Skinner J, Noble PC. Understanding why metal-on-metal hip arthroplasties fail. J Bone Jt Surgery-American. 2012;94:e22–1–10.

27. Mabilleau G, Kwon Y-M, Pandit H, Murray DW, Sabokbar A. Metal-on-metal hip resurfacing arthroplasty: a review of periprosthetic biological reactions. Acta Orthop. 2008;79:734–47.

28. Athanasou NA. The pathobiology and pathology of aseptic implant failure. Bone Joint Res. 2016;5:162–8.

29. Kwon Y-M, Thomas P, Summer B, Pandit H, Taylor A, Beard D, et al. Lymphocyte proliferation responses in patients with pseudotumors following metal-on-metal hip resurfacing arthroplasty. J Orthop Res. 2009;28:444–50.

30. Lainiala O, Reito A, Elo P, Pajamäki J, Puolakka T, Eskelinen A. Revision of metal-on-metal hip prostheses results in marked reduction of blood cobalt and chromium ion concentrations. Clin Orthop Relat Res. 2015:1–9.

31. MHRA. Medicines and Healthcare Products Regulatory Agency. Medical device alert: all metal-on-metal (MoM) hip replacements (MDA/2010/033). 2010. https://webarchive.nationalarchives.gov.uk/20101116163355/, http://www.mhra.gov.uk/home/idcplg?IdcService=GET_FILE&dDocName=CON079162&RevisionSelectionMethod=LatestReleased. Accessed 4 Jan 2018.

32. DePuy Orthopaedics. Urgent field safety notice. DePuy ASR™ articular surface replacement and ASR™ XL acetabular system recall. 2010. https://www.depuysynthes.com/asrrecall/sites/default/files/DPYUS1%20Recall%20Notice.pdf. Accessed 4 Jan 2018.

33. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am. 1969;51:737–55.

34. Fitzpatrick R, Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J. Bone Joint Surg. 1996;78:185–90.

35. Hart AJ, Sabah SA, Bandi AS, Maggiore P, Tarassoli P, Sampson B, et al. Sensitivity and specificity of blood cobalt and chromium metal ions for predicting failure of metal-on-metal hip replacement. Bone Joint J. 2011; 93(10):1308–13.

36. Bills PJ, Racasan R, Underwood RJ, Cann P, Skinner J, Hart AJ, et al. Volumetric wear assessment of retrieved metal-on-metal hip prostheses and the impact of measurement uncertainty. Wear. 2012;274–275:212–9.

37. Davies a P, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. J bone Jt surgeryAmerican. 2005;87:18–27.

38. Bonnini S, Corain L, Marozzi M, Salmaso L. Nonparametric hypothesis testing: rank and permutation methods with applications in R. Hoboken, New Jersey: Wiley; 2014.

39. Sidaginamale RP, Joyce TJ, Lord JK, Jefferson R, Blain PG, Nargol AVF, et al. Blood metal ion testing is an effective screening tool to identify poorly performing metal-on-metal bearing surfaces. Bone Joint Res. 2013;2:84–95.

40. Hart AJ, Sabah SA, Sampson B, Skinner JA, Powell JJ, Palla L, et al. Surveillance of patients with metal-on-metal hip resurfacing and Total hip prostheses: a prospective cohort study to investigate the relationship between blood metal ion levels and implant failure. J Bone Joint Surg Am. 2014;96:1091–9.

41. Kwon Y-M, Ostlere SJ, McLardy-Smith P, Athanasou NA, Gill HS, Murray DW. "Asymptomatic" pseudotumors after metal-on-metal hip resurfacing arthroplasty: prevalence and metal ion study. J Arthroplast. 2011;26:511–8.

42. Madanat R, Hussey DK, Donahue GS, Potter HG, Wallace R, Bragdon CR, et al. The symmetry of adverse local tissue reactions in patients with bilateral simultaneous and sequential ASR hip replacement. J Arthroplast. 2015;30:1794–8.

43. Pandit H, Vlychou M, Whitwell D, Crook D, Luqmani R, Ostlere S, et al. Necrotic granulomatous pseudotumours in bilateral resurfacing hip arthoplasties: evidence for a type IV immune response. Virchows Arch. 2008;453:529.

44. Uchihara Y, Grammatopoulos G, Munemoto M, Matharu G, Inagaki Y, Pandit H, et al. Implant failure in bilateral metal-on-metal hip resurfacing arthroplasties: a clinical and pathological study. J Mater Sci Mater Med. 2018;29:28.

45. Mahendra G, Pandit H, Kliskey K, Murray D, Gill HS, Athanasou N. Necrotic and inflammatory changes in metal-on-metal resurfacing hip arthroplasties: relation to implant failure and pseudotumor formation. Acta Orthop. 2009;80:653–9.

46. Gallo J, Vaculova J, Goodman SB, Konttinen YT, Thyssen JP. Contributions of human tissue analysis to understanding the mechanisms of loosening and osteolysis in total hip replacement. Acta Biomater. 2014;10:2354–66.

47. Kwon Y-M, Xia Z, Glyn-Jones S, Beard D, Gill HS, Murray DW. Dose-dependent cytotoxicity of clinically relevant cobalt nanoparticles and ions on macrophages in vitro. Biomed Mater. 2009;4:025018.

48. Lehtovirta L, Reito A, Parkkinen J, Hothi H, Henckel J, Hart A, et al. Analysis of bearing wear, whole blood and synovial fluid metal ion concentrations and histopathological findings in patients with failed ASR hip resurfacings. BMC Musculoskelet Disord. 2017;18:523.

49. Matthies AK, Racasan R, Bills P, Blunt L, Cro S, Panagiotidou A, et al. Material loss at the taper junction of retrieved large head metal-on-metal total hip replacements. J Orthop Res. 2013;31:1677–85.